DOCUMENT RESUME

ED 181 525 CS 502 803

TITLE	Status Report on Speech Research: A Report on the Status and Progress of Studies on the Nature of Speech, Instrumentation for Its Investigation, and Practical Applications, July 1 - December 31, 1979.
INSTITUTION	Haskins Labs., New Haven, Conn.
REPORT NO	SR-59/60(1979)
PUB DATE	79
NOTE	276p.
EDRS PRICE	MF01/PC12 Plus Postage.
DESCRIPTORS	*Articulation (Speech): Auditory Perception: Beginning Reading: Consonants: *Language Research: Languages: Linguistic Performance: Linguistics: Memory: Physiology: *Pronunciation: *Speech Communication: Speech Skills: Vowels
IDENTIFIERS	Communication Research: *Oral Language

ABSTRACT

Thirteen manuscripts and extended reports on the nature of speech, instrumentation for its investigation, and practical applications of research are presented in this status report on speech research for the period of July-December 1979. The topics dealt with include speech perception, children's memory for sentences and word strings in relation to reading ability, influence of preceding fricative on stop consonant perception, electromyographic study of the jaw muscles during speech, laryngeal activity in Swedish voiceless obstruent clusters, laryngeal adjustments in the production of consonant clusters and geminates in American English, lexical tone and sentence prosody in Thai, influence of vocalic environment on perception of silence in speech, stimulus dominance and ear dominance in fused dichotic speech and nonspeech stimuli, use of feedback in established and developing speech, and brain lateralization in two-, three-, and four-month-old children for phonetic and musical timbre discriminations under memory load. (FL)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED181525

U.S. DEPARTMENT OF HEALTH.
EDUCATION & WELFARE
NATIONAL INSTITUTE OF
EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

SR-59/60 (1979)

Status Report on

SPEECH RESEARCH

A Report on
the Status and Progress of Studies on
the Nature of Speech, Instrumentation
for its Investigation, and Practical
Applications

1 July - 31 December 1979

Haskins Laboratories
270 Crown Street
New Haven, Conn. 06510

Distribution of this document is unlimited.

(This document contains no information not freely available to the general public. Haskins Laboratories distributes it primarily for library use. Copies are available from the National Technical Information Service or the ERIC Document Reproduction Service. See the Appendix for order numbers of previous Status Reports.)

CS502823

ACKNOWLEDGMENTS

The research reported here was made possible in part by support from the following sources:

National Institute of Child Health and Human Development
Grant HD-01994

National Institute of Child Health and Human Development
Contract N01-HD-1-2420

National Institutes of Health
Biomedical Research Support Grant RR-05596

National Science Foundation
Grant BNS76-82023
Grant MCS76-81034
Grant MCS79-16177

National Institute of Neurological and Communicative Disorders and Stroke
Grant NS13870
Grant NS13617

National Institute of Arthritis, Metabolism, and Digestive Diseases
Grant AM25814

HASKINS LABORATORIES

Personnel in Speech Research

Alvin M. Liberman,* President and Research Director
Franklin S. Cooper,* Associate Research Director
Patrick W. Nye, Associate Research Director
Raymond C. Huey, Treasurer
Alice Dadourian, Secretary

Investigators

Arthur S. Abramson*
Thomas Baer
Fredericka Bell-Berti+
Catherine Best+
Gloria J. Borden*
Guy Carden*
Robert Crowder*
William Ewan*
Carol A. Fowler*
Jane H. Gaitenby
Katherine S. Harris*
Alice Healy*
Leonard Katz*
Scott Kelso
Andrea G. Levitt*
Isabelle Y. Liberman*
Leigh Lisker*
Anders Löfqvist[2]
Virginia Mann+
Charles Marshall
Ignatius G. Mattingly*
Nancy McGarr*
Lawrence J. Raphael*
Bruno H. Repp
Philip E. Rubin
Donald P. Shankweiler*
Michael Studdert-Kennedy*
Michael T. Turvey*
Robert Verbrugge*
Hirohide Yoshioka[1]

Technical and Support Staff

Eric L. Andreasson
Elizabeth P. Clark
Donald Hailey
Terry Halwes
Sabina D. Koroluk
Agnes M. McKeon
Nancy O'Brien
William P. Scully
Richard S. Sharkany
Leonard Szubowicz
Edward R. Wiley
David Zeichner

Students*

David Dechovitz
Laurel Dent
Laurie Feldman
Hollis Fitch
Carole E. Gelfer
David Goodman
Janette Henderson
Charles Hoequist
Kenneth Holt
Robert Katz
Morey J. Kitzman
Peter Kugler
Roland Mandler
Suzi Pollack
Patti Jo Price
Brad Rakerd
Arnold Shapiro
Louis G. Tassinary
Janet Titchener
Emily Tobey-Cullen
Betty Tuller
N. S. Viswanath
Douglas Whalen

---

*Part-time
[1] Visiting from University of Tokyo, Japan
[2] Visiting from Lund University, Sweden
+NIH Research Fellow

CONTENTS

I. Manuscripts and Extended Reports

Speech Perception--Michael Studdert-Kennedy 1

Children's Memory for Sentences and Word Strings in Relation to Reading Ability--Virginia A. Mann, Isabelle Y. Liberman and Donald Shankweiler 23

Effects of Vocalic Formant Transitions and Vowel Quality on the English [s]-[š] Boundary--D. H. Whalen 35

Influence of Vocalic Context on Perception of the [ʃ]-[s] Distinction: I. Temporal Factors--Virginia A. Mann and Bruno H. Repp 49

Influence of Preceding Fricative on Stop Consonant Perception--Virginia A. Mann and Bruno H. Repp 65

Electromyographic Study of the Jaw Muscles during Speech--Betty Tuller, Katherine S. Harris and Bob Gross 83

Laryngeal Activity in Swedish Voiceless Obstruent Clusters--Anders Löfqvist and Hirohide Yoshioka 103

Laryngeal Adjustments in the Production of Consonant Clusters and Geminates in American English--Hirohide Yoshioka, Anders Löfqvist and Hajime Hirose 127

Lexical Tone and Sentence Prosody in Thai--Arthur S. Abramson 153

Influence of Vocalic Environment on Perception of Silence in Speech: Amplitude Effects--Bruno H. Repp 161

Stimulus Dominance and Ear Dominance in Fused Dichotic Speech and Nonspeech Stimuli: A Replication--Bruno H. Repp 169

Use of Feedback in Established and Developing Speech--Gloria J. Borden 179

Brain Lateralization in 2-, 3-, and 4-Month-Olds for Phonetic and Musical Timbre Discriminations under Memory Load--Catherine T. Best, Harry Hoffman and Bradley B. Glanville 197

Cross-Series Adaptation Using Song and String--Robert E. Remez, James E. Cutting and Michael Studdert-Kennedy 227

Preservation of Vocal Tract Length in Speech: A Negative Finding--Betty Tuller and Hollis L. Fitch 239

Epimenides at the Computer (A review of Gödel, Escher, Bach: An Eternal Golden Braid, by Douglas R. Hofstadter)--Ignatius G. Mattingly 245

Language by Hand and by Eye (A Review of The Signs of Language, by Edward S. Klima and Ursula Bellugi)--Michael Studdert-Kennedy 251

II. Publications 265

III. Appendix: DDC and ERIC numbers (SR-21/22 - SR-57) 271

I. MANUSCRIPTS AND EXTENDED REPORTS

SPEECH PERCEPTION[*]

Michael Studdert-Kennedy+

Abstract. The paper reviews selected studies in speech perception, most of them published in the past five years. Topics include: the contributions of prosody to segmental perception, the problems of segmentation and invariance, categorical perception of speech and non-speech, the role of feature detectors, the scaling of speech sounds to an auditory-articulatory space, acoustic phonetic dependencies within the syllable, the contributions of higher order (non-phonetic) factors to the comprehension of fluent speech, and cerebral specialization. The bias of the paper is toward viewing phonetic segments as abstract processes that link sound and articulation, and that become available to the listener through specialized sensorimotor mechanisms.

The past few years of research in speech perception have been very active. The old questions are still there--What are the units? How do we segment? What are the invariants?--but some old answers have turned out to be wrong and some new ones are beginning to emerge. The intricate articulatory and acoustic structure of the syllable is still at the center of the maze, but other sources of information for the listener--prosody, syntax, semantics--have begun to receive experimental attention: Studies of fluent speech are taking their place beside the established methods of syllable analysis and synthesis. Theory has dropped into the background (or perhaps the back room) and no one seems very eager to argue the merits of analysis-by-synthesis or the "motor theory" any more. Certainly, theory continues to guide research, but a refreshing atheoretical breeze has been blowing in from artificial speech understanding research (Klatt, 1977, 1980) and from developmental psychology (Aslin & Pisoni, Note 1). In the latter regard, I shall not have much to say directly about infant speech perception, but much of what I have to say will bear on it. The infant is a listener, a very attentive one, because by learning to listen it learns to speak. In my opinion, only by carefully tracking the infant through its first two years of life shall we

---

[*] A revision of an invited Status Report on Speech Perception, prepared for pre-conference distribution to participants in the IXth International Congress on Phonetic Sciences, held in Copenhagen, August 6-9, 1979. This version will be published with the other Status Reports of the Congress in Language and Speech, Vol. 23, Part 1, Jan.-March, 1980.

+Also, Queens College and Graduate Center of the City University of New York.

Acknowledgment: I thank Alvin Liberman, David Pisoni, Terry Halwes and, especially, Bruno Repp for their careful comments on the text. Preparation of this paper was supported in part by NICHD Grant HD-01994.

[HASKINS LABORATORIES: Status Report on Speech Research SR-59/60 (1979)]

come to understand adult speech perception and, in particular, how speaking and listening establish their links at the basis of the language system. This said, let us begin, as infants do, with prosody.

Prosody

Prosody refers to the melody, rhythm, rate, amplitude, quality and temporal organization of speech. There has been an upsurge of interest in these factors in recent years, partly because they seem to hold a key to improved speech synthesis, partly because prosodic contributions to speech perception have been unjustly neglected (Cohen & Nooteboom, 1975; Nooteboom, Brokx, & de Rooij, 1976). To say that prosody "contributes" to speech perception may seem to imply that speech perception is confined to segmental processes, of which prosody is a mere subsidiary, conveying no distinctive information of its own. This, of course, is false. Prosody carries much of that important indexical information (Abercrombie, 1967) without which, if it is dark, you don't know who is talking to you or whether the talker means what he says. However, it is with the adjutant functions--contributions to segmental perception--that I am concerned here.

One prosodic function is to maintain a coherent auditory signal. Darwin (1975) asked listeners to shadow a sentence on one ear, while a competing sentence was led into the other. At some arbitrary point, prosodic contours were suddenly switched across ears, while syntactic and semantic sequences were maintained. Prosodic continuity then often overrode syntax, semantics and ear of entry, leading to the intrusion of words from the supposedly unattended ear. Evidently, listeners were tracking the prosodic contour, a process that Nooteboom et al. (1976) suggest may be necessary to maintain "perceptual integrity."

What physical dimensions of the signal sustain this integrity? Rate is probably not important, because quite sharp rate variations are regularly used to convey syntactic information (e.g., Klatt, 1976). Of course, rate can affect segmental classification (Ainsworth, 1972), but listeners adjust rapidly, within less than a second (Fujisaki, Nakamura, & Imoto, 1975; Summerfield, 1975; Nooteboom et al., 1976). Amplitude changes, within limits, are also probably of little importance (Darwin & Bethell-Fox, 1977). In fact, the principal determinants of prosodic continuity seem to be fundamental frequency (F_0) and spectrum: Nooteboom et al. (1976) showed that, when pitches, alternating over a 2-6 Hz range, are imposed on a sequence of three vowels, repeated at intervals of less than 150 msec, the vowels split into two streams, as though from two speakers. The effect is reduced, if the vowels are granted a degree of spectral continuity by being placed into consonantal context. This work, taken with similar studies by Dorman, Cutting, and Raphael (1975) and by Darwin and Bethell-Fox (1977), leads to the conclusion that continuity of both formant structure and F_0 underlies the perceptual integrity of running speech.

A second prosodic function is to facilitate phrasal grouping. Here the main variables seem to be F_0 and segment duration. Several studies have documented syntactic control of timing and segment duration in production (e.g., Cooper, 1976; Klatt, 1976). Klatt and Cooper (1975) show, further, that listeners expect segment duration to vary with the syntactic position of a word in a sentence. For example, they judge lengthened syllables to be more natural at the end of a clause than at the beginning or middle. Similarly, Nooteboom et al. (1976) report that listeners judge a vowel of a particular length to be shorter if it occurs at the end of a word than if it occurs at the beginning. Presumably, such observations reflect listeners' habitual use of phrase-lengthening as an aid to parsing.

The role of F_0 has been more extensively studied. For example, Collier and 't Hart (1975) constructed synthetic utterances consisting of 13 or 15 200-msec steady-state, vowel-like "syllables," separated by 50-msec silent intervals. They imposed ten theoretically derived F_0 contours ('t Hart & Cohen, 1973) on these syllables, deploying characteristic "continuation rises" and "non-final falls" to delimit the ends and the beginnings, respectively, of possible syntactic constituents. Finally, following Svensson (1974) and Kozhevnikov and Chistovich (1965), they asked listeners to write down syntactically acceptable sentences to match each contour in number of syllables, location of stresses and overall intonation. Of the resulting sentences, 72% matched the predicted syntactic structures. Since two hypotheses were under test here--both the correctness of the theoretically derived contours and the listeners' capacity to infer syntactic structure from intonation--this is a remarkably high score.

Finally, a third perceptual function of prosody has aroused a great deal of interest in recent years. This is a function--nobody knows what it is--supposedly fulfilled by rhythm. Martin (1972) wrote a persuasive paper in which he argued that speaking involves more than a simple concatenation of motor elements: Like other motor behaviors speech is compelled, by natural constraints on the relative timing of components, to be rhythmic. Moreover, some components (syllables) are "accented," and these are predictable: Accent level (or stress) covaries with timing and the main accents are equidistant (i.e., isochronous). Finally, since "...speaking and listening are dynamically coupled rhythmic activities..." (p. 489), listeners can predict the main stresses and can use that fact to "cycle" their attention, saving it, as it were, for the more important words.

There is, in fact, evidence from phoneme-monitoring experiments that reaction time (RT) is shorter to initial phonemes in stressed words than in unstressed (Shields, McHugh, & Martin, 1974). This is apparently not due to the greater energy of the stressed words, since, if the words are presented in isolation, no RT difference appears (Shields et al., 1974). Moreover, Cutler (1976) has found that the RT difference holds, even if stress, or the lack of it, is merely "predicted" by prior prosodic contour and if the actual target is acoustically identical in both conditions. Cutler and Foss (1977) demonstrate, further, that the RT advantage is not due to syntactic form class, since it is found for stressed function words as well as for stressed content words. They conclude that the reduced reaction time may reflect heightened attention to the semantic focus of a sentence, and they cite unpublished evidence from Allen and O'Shaugnessy that "...reliable correlates of semantic focus are to be found in the fundamental frequency contour" (p. 10).

By this last point Cutler and Foss seem to be cutting themselves free from Martin's (1972) claim for isochrony, whether wisely or not remains to be seen. Lehiste (1977) has recently reopened the isochrony issue in a paper summarizing much of her research on the topic. She concludes that although isochrony is "primarily a perceptual phenomenon" (p. 253), it does have some basis in production and is therefore available for communicative use. Lehiste shows that English interstress intervals are often lengthened to signal a syntactic boundary.

Isochrony has also come under experimental scrutiny. Morton, Marcus, and Frankish (1976), recording a list of spoken digits for experimental use, discovered that acoustically (onset to onset) isochronous sequences sounded anisochronous. Moreover, listeners, asked to adjust a sequence to perceptual isochrony, made it acoustically anisochronous. Morton et al. (1976) coined the term "perceptual centers" ("P-centers") to refer to those points in a sequence of words that are equidistant when the words sound isochronous. But they were unable to locate the points or specify their acoustic correlates. Surprisingly, the P-center does not correspond to any obvious acoustic marker, such as sound onset, vowel onset or syllable peak. However, Fowler (1979) has recently discovered that "...when asked to produce isochronous sequences, talkers generate precisely the acoustic anisochronies that listeners require in order to hear a sequence as isochronous." The acoustic anisochronies apparently arise because the articulatory onsets of words beginning with sounds from different manner classes have acoustic consequences at different relative points in time. From a review of her own and related studies (e.g., Allen, 1972; Lindblom & Rapp, 1973), Fowler concludes that "...listeners judge isochrony based on acoustic information about articulatory timing rather than on some articulation-free acoustic basis." Finally, although this work seems to be a thread that might unravel isochrony, Fowler is cautious in her claims. Most of the relevant experimental studies have used monosyllables and artificially repetitive utterances. What inroads this approach can make into the apparent isochrony of phonetically heterogeneous running speech remains to be seen.

Segmentation and invariance

We turn now from the broad questions of prosody to the narrower puzzle of the syllable on which the prosody is carried. In what follows, I assume (with most other investigators) that our task is to understand the process by which phonemes or features are extracted from the signal. Let us begin with a question raised by Myers, Zhukova, Chistovich, and Mushnikov (1975): Is segmentation an auditory process, preceding phonetic classification, or an automatic consequence of classification itself? Several studies from the Pavlov Institute in Leningrad speak to the question. Chistovich, Fyodorova, Lissenko, and Zhukova (1975) showed that a sudden amplitude drop, roughly in the middle of a 460-msec steady-state vowel, caused listeners to hear either two vowels or a VCV sequence, depending on the magnitude and rate of the amplitude decrease. Subsequently, Myers et al. (1975) used an ingenious dichotic technique to suggest that such amplitude decreases are registered by the peripheral auditory system; they inferred that, since classification is presumably central, segmentation must precede classification. Finally, Zhukov, Zhukova, and Chistovich (1974) reported on the use of a similar technique to study the effects of spectral variation at segment boundaries. The investigators presented a time-varying value of F2 (roughly 2200 to 800 Hz over 200 msec) to one ear, steady-state values of F1 and F3 to the other. The latter were interrupted by a 12-15 msec pause, of which the position could be set by the subject so as to vary the fused percept from hard to soft [r], that is, from [iru] to [ir'u]. Subjects reliably set the pause so that its endpoint coincided with an F2 value of roughly 1600 Hz. Since this value is close to that of the hard-soft boundary previously determined for the steady-state isolated consonants [s] and [s'], the authors infer that listeners were also judging the soft consonant [r'] by its F2 value at onset. They conclude that "the auditory system interprets the acoustic flow as a sequence of time segments between instants of variation" (p. 237), and that it derives consonantal information by sampling formant frequencies at these instants.

However, this conclusion does not seem to be forced by the data. On the one hand, the presumed peripheral segment boundary, determined by a sharp amplitude drop, seems to have something in common with the boundary proposed by certain automatic recognition procedures for isolating syllables rather than phonemes (e.g., Mermelstein, 1975). On the other hand, an invariant formant onset is not incompatible with the use of formant movement into the following vowel as a consonantal cue (see Dorman, Studdert-Kennedy, & Raphael, 1977). My inclination therefore is to suppose that the preliminary auditory segmentation (if any) is syllabic rather than phonemic, and that within-syllable segmentation may often be synonymous with classification. I will return to this point below.

The view of the perceptual process, proposed by the Russian group, as a succession of brief time slices (rather than as the active continuous tracking suggested by studies of prosody), is close to that currently being explored by K. N. Stevens. In a succession of publications over recent years, Stevens (e.g., 1975) has elaborated on the "quantal nature of speech." He points out that, although the vocal apparatus is capable of producing a wide variety of sounds, relatively few are actually used in the languages of the world. He attributes this restriction to a nonlinear relation between articulatory and acoustic parameters: Some articulatory configurations are acoustically stable, in the sense that small changes in articulation have little acoustic effect; others are unstable in the sense that equally small changes have a substantial effect. The universal set of phonetic features is drawn from those articulatory configurations that generate acoustically stable, invariant "properties." The properties, it should be stressed, are higher order spectral configurations, rather than isolated cues such as F2 onset frequency. To define these configurations, Stevens has largely relied on computations from a vocal tract model. Finally, to assure quantal (or categorical) perception of the invariant properties and to afford the human infant a mechanism for netting them in the speech stream, Stevens postulates a matching set of innate "property detectors."

Empirical tests of the quantal theory have been few. But a recent study of English stops (Blumstein & Stevens, 1979) is a good illustration of the approach, since it deals with a notoriously context-dependent set of sounds. The goal was to demonstrate the presence of invariant properties in the acoustic signal, sufficient for recognition by fixed templates. The first step was to record two male speakers reading random lists of the voiced stops [b d g], followed by each of five vowels [i e a o u]. Short-time spectra were then determined, integrated over a 26-msec window at onset. The spectra were used to construct, by trial and error, a template fitted to each place of articulation, such that it either correctly accepted or correctly rejected the majority of utterances. Descriptions of the templates ("diffuse-rising" for alveolar, "diffuse-falling" for labial and "compact" for velar) recall the terminology of distinctive feature theory.

In the second part of the study, a corpus of utterances was collected for classification by the templates. Six subjects (4 males, 2 females) recorded five repetitions each of the voiced and voiceless stop consonants [b d g p t k], followed by each of the vowels [a e i o u], or preceded by each of the vowels [i ε a ʌ u]. The resulting 1800 utterances were then analyzed spectrally in the same way as the original utterances, and compared with the templates. The results were: at least 80% (and often higher) correct rejection and correct acceptance for initial stops, a slightly lower performance for released final stops, although for some unreleased final stops scores dropped as low as 40%. Analysis of variance revealed significant differences in template matching performance as a function of vowel context, but performance was significantly above chance in every case. Quite similar results have been reported by Searle, Jacobson, and Rayment (1979) using a very much longer time slice (100-200 msec) and deriving their invariant patterns from a running sequence of spectra.

Where then does this leave us? Eighty percent or better is a good score--although, as A. M. Liberman has suggested to me, we might do almost as well with the binary recipe proposed by Cooper, Delattre, Liberman, Borst, and Gerstman in 1952: high burst, falling F2 transition for alveolar; low burst, falling F2 transition for velar; low burst, rising F2 transition for labial.

The question, of course, is: Is this really the way that humans do it? Dorman et al. (1977), modeling their study on the work of Fischer-Jørgensen (1972), edited release bursts and/or formant transitions out of English voiced stop consonants ([b d g]), spoken before nine different vowels. Acoustic analysis of the bursts for a given place of articulation showed them to be largely invariant (cf. Zue, 1976). However, the bursts were not invariant in their effect: For the most part, listeners only perceived the bursts correctly if their main spectral weight lay close to the main formant of the following vowel, as Stevens himself has suggested (1975, pp. 312-313). Kuhn (1975) has shown that the main vowel formant varies with the length of the cavity in front of the point of maximum tongue constriction. Since front cavity length is a function of place of articulation, an estimate of front cavity resonance is tantamount to an estimate of place of articulation. Thus, proximity on the frequency scale may facilitate perceptual integration of the burst with the vowel, enabling the listener to track the changing cavity shape characteristic of a particular place of articulation followed by a particular vowel.

Stevens (see especially, 1975) does not deny that contextually variable cues--such as formant transitions, voice onset time, vowel formant structure--can be used by the human listener. However, he regards them as "secondary," learned cues, acquired by repeated association with the "primary" invariant properties, and used only as safety devices when invariant cues fail. Given the many knotty questions concerning the possible mechanisms for extracting and interpreting these "secondary" context-dependent cues, one may wonder how an organism whose primary endowment is a set of passive templates learns to use them at all.

The question becomes even more pressing when one considers that there is no independent evidence for the existence of the hypothesized templates or property detectors. To understand this we must briefly review recent findings in the study of categorical perception.

Categorical perception

As is well known, early work with speech synthesizers showed that a useful procedure for defining the acoustic properties of a phoneme was to construct tokens of opponent categories, distinguished on a single phonological feature, by varying a single acoustic parameter along a continuum (e.g., [ba] to [da], [da] to [ta], etc.). If listeners were asked to identify these tokens, they tended to identify any particular stimulus in the same way every time they heard it: There were few ambiguous tokens. Moreover, if they were asked to discriminate between neighboring tokens, they tended to do very badly if they assigned the two tokens to the same class, very well if they assigned them to different classes--even though the acoustic distance between tokens was identical in the two cases. This phenomenon was dubbed "categorical perception" (Liberman, Harris, Hoffman, & Griffith, 1957). Although there were usually no grounds for supposing that the acoustic variations along synthetic continua mimicked the intrinsic allophonic variations of natural speech, categorical perception in the laboratory was taken to reflect a necessary aspect of normal speech perception, namely, the rapid transfer of speech sounds into a phonetic or phonological code. The phenomenon was also believed by some people, including me, to be peculiar to speech (Studdert-Kennedy, Liberman, Harris, & Cooper, 1970).

However, we now know that categorical perception, as observed in the laboratory, is neither peculiar nor necessary to speech. Demonstrations that it is not peculiar we owe to Cutting and Rosner (1974) (rise-time at the onset of sawtooth waves, analogous to a fricative-affricate series); to Miller, Wier, Pastore, Kelly, and Dooling (1976) (noise-buzz sequences analogous to the aspiration-voice sequences of a voice onset time (VOT) series); to Pisoni (1977) (relative onset time of two tones); and to Pastore, Ahroon, Baffuto, Friedman, Puleo, and Fink (1977). These last investigators extended their work into vision, demonstrating categorical perception of critical flicker, with a sharp boundary at the flicker-fusion threshold. They also induced clearly categorical perception of a sine-wave intensity series by providing listeners with a constant-reference tone, or "pedestal," at the center of the series. Pastore et al. (1977) conclude that a continuum may be categorically divided either by a sensory threshold (as in flicker-fusion) or by an internal reference (as in the intensity series). Presumably, the portion of the signal with the earlier onset serves as a reference in a VOT series, while in a place of articulation series, cued by direction and extent of formant transitions, a reference is provided by the fixed vowel. If this last point is correct, we perceive a place series categorically precisely because the consonants are judged relationally rather than absolutely—an interpretation not compatible with the notion of invariant property detectors.

Just how an internal reference suppresses discrimination within categories is not clear, but the results of Carney, Widin, and Viemeister (1977) suggest that it may simply serve to divert the listener's attention from other stimuli in the series. To Carney et al. (1977) (see also Pisoni & Lazarus, 1974; Samuel, 1977) we owe the demonstration that a VOT continuum need not be perceived categorically. Each of their subjects displayed good within-category discrimination after moderate training on a bilabial VOT continuum. Indeed, discrimination was so good that subjects were able to shift category boundaries on request and assign consistent labels to arbitrary subsets of the stimuli. The outcome suggests that "...utilization of acoustic differences between speech stimuli may be determined primarily by attentional factors, ...distinct from the perceptual capacities of the organism" (Carney et al., p. 969).

This is precisely what is suggested by the numerous instances in which speakers of different languages perceive an acoustic continuum in different ways. (For a thorough review, see Strange & Jenkins, 1978.) For example, while American English speakers perceive an [r] to [l] continuum categorically, Japanese speakers do not (Miyawaki, Strange, Verbrugge, Liberman, Jenkins, & Fujimura, 1975). For another example, not only do Spanish and American English speakers place their category boundaries at different points along the VOT continuum (Abramson & Lisker, 1973; Williams, 1978), but also Spanish-English bilinguals can be induced to shift their boundaries by a shift in language set within a single test (Elman, Diehl, & Buchwald, 1977). Not unrelated, perhaps, is the recent demonstration by Ganong (1978) that listeners have a bias for words over nonwords: Offered a continuum of which one end is a word (e.g., [gɪft]) and the other not (e.g., [kɪft]), they shift their normal boundary away from the word, thus increasing the number of words they hear.

Presumably there are limits to this sort of thing. With adequate synthesis, the range of uncertainty must be limited and we may still use synthetic continua to assess "the auditory tolerance of phonological categories" (Brady & Darwin, 1978, p. 1556)--precisely the use for which they were first designed over twenty-five years ago.

Feature or property detectors

The demonstration that listeners can be trained to hear a supposedly categorical continuum noncategorically undercuts the original evidence for acoustic feature, or property, detectors in speech perception, namely, categorical perception itself. Moreover, it throws into doubt the interpretation of a substantial body of work on selective adaptation of speech sounds that has appeared in the past five years.

The series began with a paper by Eimas and Corbit (1973). They asked listeners to categorize members of a synthetic voice onset time (VOT) continuum (Lisker & Abramson, 1964) and demonstrated that the perceptual boundary between voiced and voiceless categories along that continuum was shifted by repeated exposure to (that is, adaptation with) either of the endpoint stimuli: There was a decrease in the frequency with which stimuli close to the original boundary were assigned to the adapted category and a consequent shift of the boundary toward the adapting stimulus. Since the effect could be obtained on a labial VOT continuum after adaptation with a syllable drawn from an alveolar VOT continuum, and vice versa, adaptation was clearly neither of the syllable as a whole nor of the unanalyzed phoneme, but of a feature within the syllable. Eimas and Corbit therefore termed the adaptation "selective" and attributed their results to the fatigue of specialized detectors and to the relative "sensitization" of opponent detectors. Subsequent studies replicated the results for VOT and extended them to other feature oppositions, such as place and manner of articulation. These studies have been reviewed by Cooper (1975), Ades (1976), and Eimas and Miller (1978).

Unfortunately, there are many grounds for doubting the opponent detector model. First, as already remarked, is the demonstration that listeners can be trained to discriminate at least some speech continua within categories. Second, the model lacks behavioral or neurological motivation. For, while the facts of additive color mixture make an opponent detector account of aftereffects entirely plausible, the facts of laryngeal timing or spectral scatter at stop consonant onset certainly do not. Third, the hypothesis is rendered implausible by dozens of reports of contextual effects: Adaptation of consonantal features is apparently specific to following vowel, to syllable position, to syllable structure (Hall & Blumstein, 1978) and even to fundamental frequency (Ades, 1977). As Simon and Studdert-Kennedy (1978) remark, "...the theoretical utility of selectively tuned feature detectors goes down as the number of contexts to which they must be tuned goes up." Moreover, the degree of adaptation varies quite generally with the acoustic distance between adaptor and test syllables, an effect typical of psychophysical contrast studies. In fact, Simon and Studdert-Kennedy (1978), drawing on their own work and that of Sawusch (1977), marshal evidence to show that selective adaptation along speech continua reflects a combination of peripheral auditory fatigue and central auditory contrast. They do not deny that selective adaptation has possible fruitful use in isolating functional channels of analysis. But if their argument is correct, we now have no evidence at all for specialized detector mechanisms tuned to the acoustic correlates of abstract linguistic features.

Scaling studies and feature interactions

This conclusion sits nicely with the results of many studies in which phoneme confusions or similarity judgments have been used to characterize the psychological representation of speech sounds. Although results vary widely with experimental method (van den Broecke, 1976), these studies typically find that vowels (e.g., Terbeek, 1977) and consonants (e.g., Singh, Woods, & Becker, 1972) fall readily into low-confusion/high-similarity groups isomorphic with some standard phonological feature set. However, as Goldstein (1977) has pointed out, relations within these feature groups are usually not random. Rather, the psychological space is structured in such a way as to suggest a continuous auditory representation within feature groups. Presumably, since the continuous auditory representation derives from an acoustic structure shaped by articulation, we could describe an analogous articulatory space by scaling articulatory errors. It was Goldstein's (1977) insight to hypothesize that the variance common to the auditory and articulatory spaces would then prove to be categorical. His study—too complicated for summary here—largely supported that hypothesis. We may fairly conclude that our models of perception should allow for continuous auditory and articulatory representations from which categories can only be derived by some abstract metric common to both.

The idea that speech sounds (perhaps unsegmented syllables) may be internally represented in a continuous auditory space (at some point before classification) is compatible with the repeated finding of interaction between features during perceptual processing (e.g., Sawusch & Pisoni, 1974; Miller, 1977). There is, in fact, no good reason to refer to these auditory processes as "featural" at all (Parker, 1977). Repp (1977) and Oden and Massaro (1978), for example, have already proposed specific models of integration based on a continuous spatial representation.

Steps toward an auditory-articulatory space

The view of speech perception that seems to be emerging from the studies we have reviewed is of an active, continuous process. We turn now to several studies of perceptual integration across the syllable which seem to call for just such an interpretation.

Perhaps the most familiar example is provided by voicing cues for stops in initial position. The concept of voice onset time (VOT) originally offered an *articulatory* account of how a range of disparate and incommensurable acoustic cues (including, as it happens, the interval between release burst and the onset of voicing) comes to signal the voiced-voiceless distinction. In fact, as Abramson (1977) has recently reminded us, VOT is simply a special case of the laryngeal timing mechanisms by which voicing distinctions are, in general, implemented.

To illustrate the underlying articulatory rationale, consider the suggestion by Stevens and Klatt (1974) that the duration of the first formant voiced transition might be a more potent cue than VOT itself. The motivation for the proposal seems to have been to coordinate the voicing cue with Stevens' hypothesized cues to place of articulation (rapid spectral scatter), and perhaps to avoid saddling the infant with a delicate timing mechanism. As it happens, Simon and Fourcin (1978) have shown that English-speaking children do not learn to use the F1 cue until they are five years old, while French-speaking children never use it at all. In any event, careful analysis by Lisker (1975) and by Summerfield and Haggard (1977) has shown that the principal first formant cue is not transition duration, but frequency at onset: the higher the frequency, the less likely is a sound to be judged voiced. Listeners apparently take a high first formant onset as a cue that the mouth was relatively wide open (and release therefore well past) when voicing began.

A less familiar set of cues to another distinction has recently been studied by Repp, Liberman, Eccardt, and Pesetsky (1978). They recorded the utterance: "Did anybody see the *gray ship*?" Then, by varying the durations of fricative noise at the onset of *ship* and of the silent interval between *gray* and *ship*, they explored the conditions under which the utterance was heard as ending with "gray chip," "great ship" or "great chip." Among their results was the finding that whether or not a syllable final stop was heard (*gray* vs. *great*) depended not only on the duration of the silence, but also on the duration of the noise following the silence. Just such an equivalence between a spectral property and silence emerges from an analysis of the trading relation between silence and formant transition in the cues for the medial [p] of [splɪt] (Liberman & Pisoni, 1977). How are we to rationalize such an equivalence? Repp et al. (1978) point out that neither a single feature detector nor a set of feature detectors, integrated by some higher level decision mechanism (as proposed by Massaro & Cohen, 1977), nor, it would seem, any purely auditory principle, can explain why such phenomenologically diverse cues can be traded off and integrated into a unitary percept.

As a final example, consider a positively Daedalian series of experiments by Bailey and Summerfield (1978). They explored the conditions under which a particular voiceless stop ([p], [t] or [k]) is perceived if a silence is introduced between [s] and a following vowel. Whether a stop is heard at all depends, of course, on the duration of the silence, but the effect of that duration itself depends on the onset frequency of F1, while the perceived place of articulation depends on the duration of the closure, on spectral properties at the offset of [s] and on the relation between those properties and the following vowel (cf. Dorman et al., 1977). Bailey and Summerfield suggest that, "...given sufficiently precise stimulus control, perceptual sensitivity could be demonstrated to every difference between two articulations" (p. 55) (cf. Haggard, Note 3). Again, the problem is to understand the principles by which such heterogeneous collections of spectral and temporal cues are combined into a percept. What rationalizes their integration?

The answer, explicitly proposed by the authors of these several studies, is that the cues are held together by their origin in the integral, articulatory gesture. We should be absolutely clear that this is not a form of motor theory. Rather, it is a description of what the perceptual system appears to do. The system follows the moment-to-moment acoustic flow, apprehending an auditory "motion picture," as it were, of the articulation, in a manner totally analogous to that by which the visual system might follow the optic flow to apprehend the articulation by reflected light rather than by radiated sound (cf. Fowler, 1979; Studdert-Kennedy, 1977).

Reading lips and reading spectrograms

The argument is clarified, and developed, in a recent study of lip reading by Summerfield (1979). Subjects were asked to write down a series of sentences spoken over an audio system, but simultaneously masked by the talker's own voice reading another text. There were three conditions of interest to the present discussion: (1) audio alone; (2) audio with full video of the speaker's face; (3) audio with a video display of the speaker's lips. Without any training, naive subjects scored 23%, 65% and 54% correct, respectively. In a second experiment, Summerfield analyzed errors made against deliberately conflicting video. He found, as did McGurk and McDonald (1976), that subjects frequently made judgments reflecting a compound between the auditory and visual information. Summerfield (as also Haggard, Note 2) points out that such instantaneous interplay between modalities seems to require a common metric by which the two streams of information can be combined. (The problem, incidentally, is quite general and may apply to any sound-producing visual event.)

It is instructive to compare the ease with which naive subjects used the visual display of face or lips with the obvious difficulty experienced by even the most skilled spectrogram reader. Cole, Rudnicky, Reddy, and Zue (1980) report a systematic study of subject VZ who has been studying acoustic phonetics for more than seven years and has logged some 2000-2500 hours reading spectrograms--perhaps as many hours as a child of two years has spent listening to speech. Despite the fact that VZ is free to use the ample context of vision (rather than the narrow window of audition) and that he reports conscious, acoustic-phonetic interpretation of visual context at least 18% of the time; despite the fact that he came to the spectrograms knowing that their visual segments were not isomorphic with phonetic segments (a crucial piece of knowledge that cannot be derived from the spectrograms themselves); despite the fact that, in the hours devoted to spectrograms, he could probably have learned to read several foreign languages with fair proficiency, VZ now transcribes spectrograms at a rate some 20 to 40 times real time (Cole, personal communication).

One is not surprised. There are, after all, biological constraints on learning (see Hinde & Stevenson-Hinde, 1973): Pigeons learn more readily to peck plastic keys for grain and to jump to avoid shock than vice versa. The visual display of talking lips and face is natural and its code is known to every speaker of a natural language, as the code of a spectrographic display is not. Watching its mother's face and listening to her speak, the infant learns to perceive articulation directly, whether by light or by sound.

Extracting information from the syllable

The primary unit of perception is evidently the unsegmented syllable (the rhythmic unit of nursery rhymes), and there is ample evidence for perceptual interaction among its components (see Studdert-Kennedy, 1976, for a review). For a recent example, Hasegawa and Daniloff (1976) synthesized two fricative continua, /s/ - /ʃ/, after two different vowels, /i/ and /u/, and found a significant shift in the phoneme boundary as a function of preceding vowel. Kunisaki and Fujisaki (1977) developed the finding by showing that contextual dependency in perception corrects for a mirror-image contextual dependency in production: Just as the frequencies of fricative poles and zeros are lower before /u/ than before /a/, so, in perception, the frequencies of the poles and zeros at the synthetic boundary between /s/ and /ʃ/ are higher before /a/ than before /u/. These results mesh neatly with our earlier conclusion that consonantal onset is judged as part of a dynamic, temporal pattern.

Just such a process has recently been shown to play an important role also in vowel perception. Strange, Jenkins, and Edman (1978) recorded tokens of /b/-vowel-/b/ syllables with ten different medial vowels, spoken by several speakers. They edited out the steady-state syllable nuclei (50% to 65% of the entire syllable, depending on the vowel) and presented various fragments of the syllables for identification. The results varied with both speaker and vowel, but overall, for three speakers of the same dialect as the listeners, error rates on the original syllables, on the syllables without their centers ("silent centers") and on the isolated centers were 4%, 10% and 18% respectively. The error rates for either the initial or the final transitions alone were approximately 60%. Evidently, the dynamic sweep of the spectral information and its temporal distribution across the syllable was the principal source of listener information in identifying these vowels, even when that portion usually said to characterize a vowel (namely, its steady state) was completely missing.

Results such as these return us to the segmentation issue. Clearly, there was little basis for peripheral segmentation in these syllables. In fact, one is tempted to suppose that listeners recognized syllables (Massaro, 1975) or perhaps "diphones" (Klatt, 1978) rather than phonemes. Mermelstein (1978) reports a subtle experiment that speaks to this issue. He varied the duration and first formant frequency of the steady-state nucleus of synthetic syllables to yield /bɛd/, /bæd/, /bɛt/, /bæt/. Notice that exactly the same acoustic information (namely, duration of the steady-state nucleus) controls both vowel and final consonant decision. Accordingly, if subjects are asked to determine duration boundaries for both consonant voicing and vowel quality as a function of F1 frequency, and if the boundaries prove to be correlated, then we can conclude that listeners made a single--presumably syllabic--decision. However, if the boundary values prove independent, we can conclude that listeners recognized phonemes rather than syllables and that they made two phonetic decisions on the basis of a single piece of acoustic information. This was, in fact, the outcome. If this is the normal mode of speech perception, it would seem that, even if syllabic segmentation is peripheral (cf. Myers et al., 1975), phonemic segmentation may be a central process consequent upon classification. Usually, this process is facilitated by auditory contrast within the syllable (cf. Bondarko, 1969).

Continuous speech

We come, finally, full circle to continuous speech with its prosody, syntax and "real world" constraints. Here, the main question is whether the perceptual processes we have been discussing up to this point have any bearing at all. Is it possible, for example, that, given the contextual aids of prosody, syntax, semantics, the listener needs no more than the "auditory contour" of a word (Nooteboom et al., 1976; cf. Morton & Long, 1976) or perhaps a few "invariant features" (Cole & Jakimik, 1978) to gain access to his lexicon?

I have no space for a full discussion of this issue (a beginning is made by Liberman & Studdert-Kennedy, 1977). But a good place to start is with a paper by Shockey and Reddy (1975) who studied speech recognition in the absence of phonological and all other higher order constraints. They recorded some fifty short utterances, spoken by native speakers of eleven different languages and presented them to four phoneticians for transcription. The transcriptions were then compared with a "target" description, determined from native speakers and spectral analysis. The average "correct" score for the four transcribers was 56% and their average agreement 50%. Comparable scores for transcription of a familiar language, without contextual or syntactic constraints, would be roughly 90%--the level reached by the three transcribers of Cole et al. (1978) in their spectrogram reading study, cited above, and, moreover, a level close to that of VZ himself when reading spectrograms. The difference of roughly 40% is evidently due to the transcribers' knowledge of the phonology of the language being transcribed.

The point of this example is that the main difference between listening to continuous speech in a familiar language and to isolated words in a foreign one may not be in the syntax, semantics or real world constraints so much as in the phonology. This is a simplification, since phonology and syntax are not independent. But it serves to emphasize that phonology makes linguistic communication possible by setting limits on how a speaker is permitted to articulate and what a listener can expect to hear (Liberman & Studdert-Kennedy, 1978). The problem of how the listener extracts and combines information from the signal to arrive at a unitary percept is, of course, exactly the same for continuous speech as for isolated words.

The function of the other higher order constraints--syntax, context, semantics--is facilitative. They serve to delimit the sampling space from which the listener's percepts may be drawn. This is well illustrated by several experiments of Cole and Jakimik (1978), using the ingenious "listening for mispronunciations" (LM) technique, devised by Cole (1973). Subjects are asked to listen to a recorded story into which mispronunciations have been systematically introduced. Their accuracy and speed of detection is then measured as a function of different variables. Mispronunciations prove to be more rapidly reported for high than for low transitional probability words (cf. Morton & Long, 1976), for words appropriate to a theme than for words inappropriate, for words implied by previous statements than for words not implied, and so on. Presumably the more rapid reports reflect the varied ways in which thresholds for words are lowered by contextual factors. Of course, the fact that listeners recover the words at all means that they can do so without a full phonetic analysis. But this should not be taken to mean that they can do so without any phonetic analysis at all.

By far the fullest and most careful account of the interactive processes of word recognition in continuous speech is offered by Marslen-Wilson (1975), (see also Marslen-Wilson & Welsh, 1978). His experimental procedure also involves mispronunciations, but the subjects' task is to shadow the text as rapidly as possible. Marslen-Wilson examines the effects of context on the frequency of fluent restorations. These restorations are often so fast that the shadower begins to say the correct word (e.g., "company") before the second syllable of the mispronounced word (e.g., "compsiny") has begun (cf. Kozhevnikov & Chistovich, 1965). Since such restorations occur only when the disrupted word is syntactically and semantically apt, it is evident that these higher order factors have facilitated recovery of the correct word. However, they cannot do so in the absence of all phonetic information. It is reassuring to read as the conclusion of a lengthy and subtle discussion of these matters: "...word-recognition in continuous speech is fundamentally data-driven, in the specific sense that the original selection of word-candidates is based on the acoustic-phonetic properties of the initial segment of the incoming word" (Marslen-Wilson & Welsh, 1978, p. 60; cf. Nakatani & Dukes, 1977).

Cerebral specialization

Nonetheless, opposition between the two modes of lexical access--holistic, from "auditory contour," analytic, from phonetic segments--should not be too sharply drawn. The work of Zaidel (1978a, 1978b) with "split-brain" patients has demonstrated that holistic access is certainly possible. The cerebral hemispheres of such patients have been surgically separated by section of the connecting pathways (corpus callosum) for relief of epileptic seizure. The separation permits an investigator to assess the linguistic capacities of each hemisphere independently. Zaidel (1978a, 1978b) has shown that the isolated right hemisphere of such a patient, though totally mute, can recognize a sizable auditory lexicon and has a rudimentary syntax sufficient for understanding phrases of up to three or four words in length. However, it is incapable of identifying nonsense syllables or of performing tasks that call for phonetic analysis, such as recognizing rhyme (cf. Levy, 1974). This phonetic deficit evidently precludes short-term verbal store, thus limiting the right hemisphere's capacity for syntactic analysis of lengthy utterances, and forces organization of language around meaning. Whether we assume a similar, subsidiary organization in the left hemisphere or some process of interhemispheric collaboration, it is clear that normal language comprehension could, at least in principle, draw on both holistic and analytic mechanisms.

At the same time, Zaidel's work provides striking support for the hypothesis, originally derived from dichotic studies, that the distinctive linguistic capacity of the left hemisphere is for phonological analysis of auditory pattern (Studdert-Kennedy & Shankweiler, 1970). Further support has come from electroencephalography (Wood, 1975) and, quite recently, from studies of the effects of electrical stimulation during craniotomy (Ojemann & Mateer, 1979). The latter work isolated, in four patients, left frontal, temporal and parietal sites, surrounding the final cortical motor pathway for speech, in which stimulation blocked both sequencing of oro-facial movements and phoneme identification.

This fascinating discovery meshes neatly with a growing body of data and theory that has sought, in recent years, to explain the well-known link between lateralizations for hand control and speech. Semmes (1968) offered a first account of the association by arguing, from a lengthy series of gunshot lesions, that the left hemisphere is focally organized for fine motor control, the right hemisphere diffusely organized for broader control. Subsequently, Kimura and her associates reported that skilled manual movements (Kimura & Archibald, 1974) and non-verbal oral movements (Mateer & Kimura, 1977) tend to be impaired in cases of non-fluent aphasia. These impairments are specifically for the sequencing of fine motor movements and are consistent with other behavioral evidence that motor control of the hands and of the speech apparatus is vested in related neural centers (Kinsbourne & Hicks, 1979). In fact, Kimura (1976) has proposed that "...the left hemisphere is particularly well adapted, not for symbolic function _per se_, but for the execution of some categories of motor activity which happened to lend themselves readily to communication" (p. 154). Among these categories we must, incidentally, include those that support the complex "phonological" and morphological processes of manual sign languages, now being discovered by the research of Klima, Bellugi and their colleagues (Klima & Bellugi, 1979).

The drift of all this work is toward a view of the left cerebral hemisphere as the locus of interrelated sensorimotor centers, essential to the development of language, whether spoken or signed. To understanding of the speech sensorimotor system perceptual studies of dichotic listening will doubtless contribute. Indeed, important dichotic studies have recently found evidence for the double dissociation of left and right hemisphere, speech and music, in infants as young as two or three months (Entus, 1977; Glanville, Best, & Levenson, 1977). However, dichotic work has not fulfilled its early promise, largely because it has proved extraordinarily difficult to partial out the complex of factors, behavioral and neurological, that determine the degree of observed ear advantage (cf. Studdert-Kennedy, 1975). For the future, we may increasingly rely on instrumental techniques for monitoring brain activity, such as the blood-flow studies of Lassen and his colleagues (Lassen, Ingvar, & Skinhoj, 1978), induced reversible lesions by focal cooling (Zaidel, 1978b), improved methods of electroencephalographic analysis, auditory evoked potentials (Molfese, Freeman, & Palermo, 1975) and, perhaps infrequently, direct brain stimulation.

REFERENCE NOTES

1. Aslin, R. N., & Pisoni, D. B. *Some developmental processes in speech perception*. Paper presented at the NICHD Conference "Child Phonology: Perception, Production and Deviation," Bethesda, Md., May 28-31, 1978.
2. Haggard, M. P. *Do we want a theory of speech perception?* Paper presented to the Research Conference on Speech-Processing Aids for the Deaf, Gallaudet College, Washington, D.C., May 23-26, 1977.

REFERENCES

Abercrombie, D. *Elements of general phonetics*. Chicago: Aldine, 1967.

Abramson, A. S. Laryngeal timing in consonant distinctions. *Phonetica*, 1977, *34*, 295-303.

Abramson, A., & Lisker, L. Voice timing perception in Spanish word-initial stops. *Journal of Phonetics*, 1973, *1*, 1-8.

Ades, A. E. Adapting the property detectors for speech perception. In R. J. Wales & E. Walker (Eds.), *New approaches to language mechanisms*. Amsterdam: North Holland, 1976.

Ades, A. E. Source assignment and feature extraction in speech. *Journal of Experimental Psychology: Human Perception and Performance*, 1977, *3*, 673-685.

Ainsworth, W. A. Duration as a cue in the recognition of synthetic vowels. *Journal of the Acoustical Society of America*, 1972, *51*, 648-651.

Allen, G. The location of rhythmic stress beats in English: An experimental study. *Language and Speech*, 1972, *15*, 72-100.

Bailey, P. J., & Summerfield, Q. Some observations on the perception of [s] + stop clusters. *Haskins Laboratories Status Report on Speech Research*, 1978, *SR-53(2)*, 25-60.

Blumstein, S. E., & Stevens, K. N. Acoustic invariance in speech production. *Journal of the Acoustical Society of America*, 1979, *66*, 1001-1017.

Bondarko, L. V. The syllable structure of speech and distinctive features of phonemes. *Phonetica*, 1969, *20*, 1-40.

Brady, S. A., & Darwin, C. J. Range effect in the perception of voicing. *Journal of the Acoustical Society of America*, 1978, *63*, 1556-1558.

Campbell, R., & Dodd, B. Hearing by eye. *Quarterly Journal of Experimental Psychology*, in press.

Carney, A. E., Widin, G. P., & Viemeister, N. F. Noncategorical perception of stop consonants differing in VOT. *Journal of the Acoustical Society of America*, 1977, *62*, 961-970.

Chistovich, L. A., Fyodorova, N. A., Lissenko, D. M., & Zhukova, M. G. Auditory segmentation of acoustic flow and its possible role in speech processing. In G. Fant & M. A. A. Tatham (Eds.), *Auditory analysis and the perception of speech*. New York: Academic Press, 1975, 221-232.

Cohen, A., & Nooteboom, S. G. (Eds.). *Structure and process in speech perception*. New York: Springer-Verlag, 1975.

Cole, R. A. Listening for mispronunciations: A measure of what we hear during speech. *Perception & Psychophysics*, 1973, *1*, 153-156.

Cole, R. A., & Jakimik, J. Understanding speech: How words are heard. In G. Underwood (Ed.), *Strategies of information processing*. New York: Academic Press, 1978, 67-116.

Cole, R. A., Rudnicky, A., Reddy, R., & Zue, V. W. Speech as patterns on paper. In R. A. Cole (Ed.), *Perception and production of fluent speech*. Hillsdale, N.J.: Erlbaum, 1980.

Cole, R. A., & Scott, B. Perception of temporal order in speech: The role of vowel transitions. *Canadian Journal of Psychology*, 1973, *27*, 441-449.

Collier, R., & 't Hart, J. The role of intonation in speech perception. In A. Cohen & S. E. G. Nooteboom (Eds.), *Structure and process in speech perception*. New York: Springer-Verlag, 1975, 107-123.

Cooper, F. S., Delattre, P. C., Liberman, A. M., Borst, J. M., & Gerstman, L. J. Some experiments on the perception of synthetic speech sounds. *Journal of the Acoustical Society of America*, 1952, *24*, 597-606.

Cooper, W. E. Selective adaptation of speech. In F. Restle, R. M. Shiffrin, N. J. Castellan, H. Lindman, & D. B. Pisoni (Eds.), *Cognitive theory*. Hillsdale, N.J.: Erlbaum, 1975.

Cooper, W. E. Syntactic control of timing in speech production: A study of complement clauses. *Journal of Phonetics*, 1976, *4*, 151-171.

Cutler, A. Phoneme-monitoring reaction time as a function of preceding intonation contour. *Perception & Psychophysics*, 1976, *20*, 55-60.

Cutler, A., & Foss, D. J. On the role of sentence stress in sentence processing. *Language and Speech*, 1977, *20*, 1-10.

Cutting, J. E., & Rosner, B. S. Categories and boundaries in speech and music. *Perception & Psychophysics*, 1974, *16*, 564-570.

Darwin, C. J. On the dynamic use of prosody in speech perception. In A. Cohen & S. E. G. Nooteboom (Eds.), *Structure and process in speech perception*. New York: Springer-Verlag, 1975.

Darwin, C. J., & Bethell-Fox, C. E. Pitch continuity and speech source attribution. *Journal of Experimental Psychology: Human Perception and Performance*, 1977, *3*, 665-672.

Dorman, M. F., Cutting, J. E., & Raphael, L. Perception of temporal order in vowel sequences with and without formant transitions. *Journal of Experimental Psychology: Human Perception and Performance*, 1975, *1*, 121-129.

Dorman, M. F., Studdert-Kennedy, M., & Raphael, L. J. Stop consonant recognition: Release bursts and formant transitions as functionally equivalent, context-dependent cues. *Perception & Psychophysics*, 1977, *22*, 109-122.

Eimas, P. D., & Corbit, J. D. Selective adaptaion of linguistic feature detectors. *Cognitive Psychology*, 1973, *4*, 99-109.

Eimas, P. D., & Miller, J. L. Effects of selective adaptation on the perception of speech and visual patterns: Evidence for feature detectors. In R. D. Walk & H. L. Pick, Jr. (Eds.), *Perception and experience*. New York: Plenum, 1978.

Elman, J. L., Diehl, R. L., & Buchwald, S. E. Perceptual switching in bilinguals. *Journal of the Acoustical Society of America*, 1977, *62*, 971-974.

Entus, A. K. Hemispheric asymmetry in processing dichotically presented speech and nonspeech stimuli by infants. In S. J. Segalowitz & F. A. Greber (Eds.), *Language development and neurological theory*. New York: Academic Press, 1977, 64-73.

Fischer-Jørgensen, E. Tape cutting experiments with Danish stop consonants in initial position. *Annual Report* (Institute of Phonetics, University of Copenhagen), 1972, *VII*.

Fowler, C. A. "Perceptual centers" in speech production and perception. *Perception & Psychophysics*, 1979, *25*, 375-388.

Fujisaki, H., Nakamura, K., & Imoto, T. Auditory perception of duration of speech and non-speech stimuli. In G. Fant & M. A. A. Tatham (Eds.), *Auditory analysis and perception of speech*. New York: Academic Press, 1975, 197-220.

Ganong, F. A word advantage in phoneme boundary experiments. *Journal of the Acoustical Society of America*, 1978, *63*, 520. (Abstract)

Glanville, B. B., Best, C. T., & Levenson, R. A cardiac measure of asymmetries in infant auditory perception. *Developmental Psychology*, 1977, *13*, 54-59.

Goldstein, L. Categorical features in speech perception and production. *Working Papers in Phonetics* (University of California at Los Angeles), 1977, *39*. (Also to appear in V. Fromkin (Ed.), *Proceedings of the Workshop on Slips of the Tongue and Ear*, Vienna.)

Hall, L. L., & Blumstein, S. E. The effect of syllabic stress and syllabic organization on the identification of speech sounds. *Perception & Psychophysics*, 1978, *24*, 137-144.

't Hart, J., & Cohen, A. Intonation by rule: A perceptual quest. *Journal of Phonetics*, 1973, *1*, 309-327.

Hasegawa, A., & Daniloff, R. G. Effects of vowel context upon labeling the /s/ - /ʃ/ continuum. *Journal of the Acoustical Society of America*, 1976, *59*, 525. (Abstract)

Hinde, R. A., & Stevenson-Hinde, J. *Constraints on learning*. New York: Academic Press, 1973.

Kimura, D. The neural basis of language *qua* gesture. In H. Whitaker & H. A. Whitaker (Eds.), *Studies in neurolinguistics* (Vol. 3). New York: Academic Press, 1976.

Kimura, D., & Archibald, Y. Motor functions of the left hemisphere. *Brain*, 1974, *97*, 337-350.

Kinsbourne, M., & Hicks, R. E. Mapping cerebral functional space: Competition and collaboration in human performance. In M. Kinsbourne (Ed.), *Asymmetrical function of the brain*. New York: Cambridge University Press, 1979, 267-273.

Klatt, D. H. Linguistic uses of segmental duration in English: Acoustic and perceptual evidence. *Journal of the Acoustical Society of America*, 1976, *59*, 1208-1221.

Klatt, D. H. Review of the ARPA Speech Understanding Project. *Journal of the Acoustical Society of America*, 1977, *62*, 1345-1366.

Klatt, D. H. Speech perception: A model of acoustic phonetic analysis and lexical access. In R. A. Cole (Ed.), *Perception and production of fluent speech*. Hillsdale, N.J.: Erlbaum, 1980.

Klatt, D. H., & Cooper, W. E. Perception of segment duration in sentence context. In A. Cohen & S. E. G. Nooteboom (Eds.), *Structure and process in speech perception*. New York: Springer-Verlag, 1975, 69-89.

Klima, E. S., & Bellugi, U. *The signs of language*. Cambridge, Mass.: Harvard University Press, 1979.

Kozhevnikov, V. A., & Chistovich, L. A. Rech 'artikuliatsiia i vospriiatiie, Moscow-Leningrad. Translated as: *Speech Articulation and Perception*. Washington, D.C.: Clearinghouse for Federal Scientific and Technical

Information, 1965, *JPRS 30-543*.

Kuhn, G. M. On the front cavity resonance and its possible role in speech perception. *Journal of the Acoustical Society of America*, 1975, *58*, 428-433.

Kunisaki, O., & Fujisaki, H. On the influence of context upon perception of voiceless fricative consonants. *Annual Bulletin* (Research Institute of Logopedics and Phoniatrics, University of Tokyo), 1977, *11*, 85-91.

Lassen, N. A., Ingvar, D. H., & Skinhoj, E. Brain function and blood flow. *Scientific American*, 1978, *239*, 62-71.

Lehiste, I. Isochrony reconsidered. *Journal of Phonetics*, 1977, *5*, 253-264.

Levy, J. Psychobiological implications of bilateral asymmetry. In S. J. Dimond & J. G. Beaumont (Eds.), *Hemisphere function in the human brain*. London: Elek, 1974.

Liberman, A. M., Harris, K. S., Hoffman, H. S., & Griffith, B. C. The discrimination of speech sounds within and across phoneme boundaries. *Journal of Experimental Psychology*, 1957, *53*, 358-368.

Liberman, A. M., & Pisoni, D. B. Evidence for a special speech-perceiving subsystem in the human. In T. H. Bullock (Ed.), *Recognition of complex acoustic signals*. Berlin: Dahlem Konferenzen, 1977, 59-76.

Liberman, A. M., & Studdert-Kennedy, M. Phonetic perception. In R. Held, H. Leibowitz, & H.-L. Teuber (Eds.), *Handbook of sensory physiology, Vol. VIII: Perception*. Heidelberg: Springer-Verlag, 1978, 143-178.

Lindblom, B., & Rapp, K. Some temporal regularities of spoken Swedish. *Papers from the Institute of Linguistics* (University of Stockholm), 1973, *21*, 1-59.

Lisker, L. Is it VOT or a first-formant transition detector? *Journal of the Acoustical Society of America*, 1975, *57*, 1547-1551.

Lisker, L., & Abramson, A. A cross-language study of voicing in initial stops: Acoustical measurements. *Word*, 1964, *20*, 384-422.

Marslen-Wilson, W. D. Sentence perception as an interactive parallel process. *Science*, 1975, *189*, 226-228.

Marslen-Wilson, W. D., & Welch, A. Processing interactions and lexical access during word recognition in continuous speech. *Cognitive Psychology*, 1978, *10*, 29-63.

Martin, J. G. Rhythmic (hierarchical) versus serial structure in speech and other behavior. *Psychological Review*, 1972, *79*, 487-509.

Massaro, D. W. Preperceptual images, processing time, and perceptual units in speech perception. In D. W. Massaro (Ed.), *Understanding language*. New York: Academic Press, 1975.

Massaro, D. W., & Cohen, M. M. Voice onset time and fundamental frequency as cues to the /zi/ - /si/ distinction. *Perception & Psychophysics*, 1977, *22*, 373-382.

Mateer, C., & Kimura, D. Impairment of non-verbal oral movements in aphasia. *Brain and Language*, 1977, *4*, 262-276.

McGurk, H., & McDonald, J. Hearing lips and seeing voices. *Nature*, 1976, *264*, 746-748.

Mermelstein, P. A phonetic context-controlled strategy for segmentation and phonetic labeling of speech. *IEEE Transactions on Acoustics, Speech and Signal Processing*, 1975, *ASSP-23*, 79-82.

Mermelstein, P. On the relationship between vowel and consonant identification when cued by the same acoustic information. *Perception & Psychophysics*, 1978, *23*, 331-336.

Miller, J. L. Nonindependence of feature processing in initial consonants.

Journal of Speech and Hearing Research, 1977, 20, 519-528.

Miller, J. D., Wier, C. C., Pastore, R., Kelly, W. J., & Dooling, R. J. Discrimination and labeling of noise-buzz sequences with varying noise-lead times: An example of categorical perception. Journal of the Acoustical Society of America, 1976, 60, 410-417.

Miyawaki, K., Strange, W., Verbrugge, R., Liberman, A. M., Jenkins, J. J., & Fujimura, O. An effect of linguistic experience: The discrimination of [r] and [l] by native speakers of Japanese and English. Perception & Psychophysics, 1975, 18, 331-340.

Molfese, D. L., Freeman, R. B., & Palermo, D. S. The ontogeny of brain lateralization for speech and nonspeech stimuli. Brain and Language, 1975, 2, 356-368.

Morton, J., & Long, J. Effect of word transition probability on phoneme identification. Journal of Verbal Learning and Verbal Behavior, 1976, 15, 43-51.

Morton, J., Marcus, S., & Frankish, C. Perceptual centers (P-centers). Psychological Review, 1976, 83, 405-408.

Myers, T. F., Zhukova, M. G., Chistovich, L. A., & Mushnikov, V. N. Auditory segmentation and the method of dichotic stimulation. In G. Fant & M. A. A. Tatham (Eds.), Auditory analysis and the perception of speech. New York: Academic Press, 1975, 243-274.

Nakatani, L. H., & Dukes, K. D. Locus of segmental cues for word juncture. Journal of the Acoustical Society of America, 1977, 62, 714-719.

Nooteboom, S. G., Brokx, J. P. L., & de Rooij, J. J. Contributions of prosody to speech perception. Institute for Perception Research Preprint (Eindhoven, Netherlands), 1976. (Also to be published in W. J. M. Levelt & G. B. Flores d' Arcais (Eds.), Studies in language perception. New York: Wiley.)

Oden, G. C., & Massaro, D. W. Integration of featural information in speech perception. Psychological Review, 1978, 85, 172-191.

Ohman, S. E. G. What is it that we perceive when we perceive speech? In A. Cohen & S. E. G. Nooteboom (Eds.), Structure and process in speech perception. New York: Springer-Verlag, 1975, 36-48.

Ojemann, G., & Mateer, C. Human language cortex: Localization of memory, syntax and sequential motor-phoneme identification systems. Science, 1979, 205, 1401-1403.

Parker, F. Distinctive features and acoustic cues. Journal of the Acoustical Society of America, 1977, 62, 1051-1054.

Pastore, R. E., Ahroon, W. A., Baffuto, K. J., Friedman, C., Puleo, J. S., & Fink, E. A. Common factor model of categorical perception. Journal of Experimental Psychology: Human Perception and Performance, 1977, 3, 686-696.

Pisoni, D. B. Identification and discrimination of the relative onset of two component tones: Implications for the perception of voicing in stops. Journal of the Acoustical Society of America, 1977, 61, 1352-1361.

Pisoni, D. B., & Lazarus, J. H. Categorical and non-categorical modes of speech perception along the voicing continuum. Journal of the Acoustical Society of America, 1974, 55, 328-333.

Repp, B. H. Dichotic competition of speech sounds: The role of acoustic stimulus structure. Journal of Experimental Psychology: Human Perception and Performance, 1977, 3, 37-50.

Repp, B. H., Liberman, A. M., Eccardt, T., & Pesetsky, D. Perceptual integration of cues for stop, fricative and affricate manner. Haskins

Laboratories Status Report on Speech Research, 1978, *SR-53(2)*, 61-83.

Samuel, A. G. The effect of discrimination training on speech perception: Noncategorical perception. *Perception & Psychophysics*, 1977, *22*, 321-330.

Sawusch, J. R. Peripheral and central processing in speech perception. *Journal of the Acoustical Society of America*, 1977, *62*, 738-750.

Sawusch, J. R., & Pisoni, D. B. On the identification of place and voicing features in synthetic stop consonants. *Journal of Phonetics*, 1974, *2*, 181-194.

Searle, C. J., Jacobson, J. Z., & Rayment, S. G. Stop consonant discrimination based on human audition. *Journal of the Acoustical Society of America*, 1979, *65*, 799-809.

Semmes, J. Hemispheric specialization: A possible clue to mechanism. *Neuropsychologia*, 1968, *6*, 11-26.

Shields, J. L., McHugh, A., & Martin, J. G. Reaction time to phoneme targets as a function of rhythmic cues in continuous speech. *Journal of Experimental Psychology*, 1974, *102*, 250-255.

Shockey, L., & Reddy, R. Quantitative analysis of speech perception. In G. Fant (Ed.), *Proceedings of the Speech Communication Seminar* (Stockholm, Sweden). New York: Wiley, 1975.

Simon, C., & Fourcin, A. J. Cross-language study of speech-pattern learning. *Journal of the Acoustical Society of America*, 1978, *63*, 925-935.

Simon, H. J., & Studdert-Kennedy, M. Selective anchoring and adaptation of phonetic and nonphonetic continua. *Journal of the Acoustical Society of America*, 1978, *64*, 1338-1357.

Singh, S. Distinctive features: A measurement of consonant perception. In S. Singh (Ed.), *Measurement procedures in speech, hearing and language*. Baltimore: University Park Press, 1978, 93-155.

Singh, S., Woods, D. R., & Becker, G. M. Perceptual structure of 22 prevocalic English consonants. *Journal of the Acoustical Society of America*, 1972, *52*, 1698-1713.

Stevens, K. N. The potential role of property detectors in the perception of consonants. In G. Fant & M. A. A. Tatham (Eds.), *Auditory analysis and perception of speech*. New York: Academic Press, 1975, 303-330.

Stevens, K. N., & Klatt, D. H. Role of formant transitions in the voiced-voiceless distinction for stops. *Journal of the Acoustical Society of America*, 1974, *55*, 653-659.

Strange, W., & Jenkins, J. J. The role of linguistic experience in the perception of speech. In H. L. Pick, Jr. & R. D. Walk (Eds.), *Perception and experience*. New York: Plenum, 1978.

Strange, W., Jenkins, J. J., & Edman, T. R. Dynamic information specifies vowel identity. *Journal of the Acoustical Society of America*, 1978, *63*. (Abstract)

Strange, W., Verbrugge, R., Shankweiler, D. P., & Edman, T. R. Consonant environment specifies vowel identity. *Journal of the Acoustical Society of America*, 1976, *60*, 213-224.

Studdert-Kennedy, M. Two questions. *Brain and Language*, 1975, *2*, 123-130.

Studdert-Kennedy, M. Speech perception. In N. J. Lass (Ed.), *Contemporary issues in experimental phonetics*. New York: Academic Press, 1976, 243-293.

Studdert-Kennedy, M. Universals in phonetic structure and their role in linguistic communication. In T. H. Bullock (Ed.), *Recognition of complex acoustic signals*. Berlin: Dahlem Konferenzen, 1977, 37-48.

Studdert-Kennedy, M., Liberman, A. M., Harris, K. S., & Cooper, F. S. Motor theory of speech perception: A reply to Lane's critical review. *Psychological Review*, 1970, *77*, 234-249.

Studdert-Kennedy, M., & Shankweiler, D. P. Hemispheric specialization for speech perception. *Journal of the Acoustical Society of America*, 1970, *48*, 579-594.

Summerfield, Q. How a full account of segmental perception depends on prosody and vice versa. In A. Cohen & S. E. G. Nooteboom (Eds.), *Structure and process in speech perception*. New York: Springer-Verlag, 1975, 51-68.

Summerfield, Q. Use of visual information for phonetic perception. *Phonetica*, 1979, *36*, 314-331.

Summerfield, Q., & Haggard, M. On the dissociation of spectral and temporal cues to the voicing distinction in initial stop consonants. *Journal of the Acoustical Society of America*, 1977, *62*, 436-448.

Svensson, S. G. Prosody and grammar in speech perception. *MILUS* (University of Stockholm), 1974, *2*.

Terbeek, D. Across-language multi-dimensional scaling of study of vowel perception. *Working Papers in Phonetics* (University of California at Los Angeles), 1977, *37*.

van den Broecke, M. P. R. *Hierarchies and rank orders in distinctive features*. Utrecht: Van Gorcum-Assen., 1976.

Williams, L. The perception of stop consonant voicing by Spanish-English bilinguals. *Perception & Psychophysics*, 1977, *21*, 289-297.

Wood, C. C. Auditory and phonetic levels of processing in speech perception: Neurophysiological and information-processing analysis. *Journal of Experimental Psychology: Human Perception and Performance*, 1975, *104*, 3-20.

Zaidel, E. Lexical organization in the right hemisphere. In P. A. Buser & A. Rougeul-Buser (Eds.), *Cerebral correlates of conscious experience*. Amsterdam: Elsevier/North Holland Biomedical Press, 1978, 177-197. (a)

Zaidel, E. Concepts of cerebral dominance in the split-brain. In P. A. Buser & A. Rougeul-Buser (Eds.), *Cerebral correlates of conscious experience*. Amsterdam: Elsevier/North Holland Biomedical Press, 1978, 263-285. (b)

Zhukov, S. Ya., Zhukova, M. G., & Chistovich, L. A. Some new concepts in the auditory analysis of acoustic flow. *Soviet Physics and Acoustics*, 1974, *20*, 237-240.

Zue, V. M. Acoustic characteristics of stop consonants: A controlled study. *Lincoln Laboratory Technical Report* (Massachusetts Institute of Technology, Cambridge, Massachusetts), 1976, *523*.

CHILDREN'S MEMORY FOR SENTENCES AND WORD STRINGS IN RELATION TO READING ABILITY*

Virginia A. Mann, Isabelle Y. Liberman+ and Donald Shankweiler+

Abstract. This study explores earlier indications that ability to make effective use of speech coding in working memory is correlated with success in learning to read. A previous study of recall of letter strings in good and poor beginning readers (Liberman, Shankweiler, Liberman, Fowler, & Fischer, 1977) revealed that the performance of good readers was more severely penalized than that of poor readers when the letter names rhymed. In order to determine whether the differences in susceptibility to phonetic interference extend to materials that more closely resemble actual text, we designed an experiment to test recall of phonetically controlled sentences and word strings. As in the case of letter recall, we found an interaction between reading ability and the effect of phonetic confusability: Though good readers made fewer errors than poor readers when sentences or word strings contained no rhyming words, they made as many errors as poor readers when many rhyming words were present. In contrast to the effectiveness of manipulations of phonetic content, systematic manipulations of meaningfulness and variations in syntactic structure did not differentially affect the two reading groups. We conclude that the inferior performance of poor readers in recall of phonetically nonconfusable sentences, word strings, and letter strings reflects failure to make full use of phonetic coding in working memory.

INTRODUCTION

Much evidence suggests that adult subjects employ a phonetic representation during comprehension of both spoken and written material (see, for

---

*A revised version of this manuscript will appear in Memory & Cognition.

+Also at University of Connecticut, Storrs.

Acknowledgment: This work was partially supported by a grant from the National Institute of Child Health and Human Development (Grant HD-01994) and by a NICHD Postdoctoral Fellowship (Fellowship HD05677) to Virginia Mann. Thanks are due to Robert Katz who also aided in testing reading ability and IQ. Bruno Repp and Leonard Katz provided helpful assistance with the statistical treatment of the data. We also wish to express our appreciation for the generous cooperation of the staff of the Parker Memorial School in Tolland, Connecticut. Special thanks are due to the principal, Robert Lincoln, the three second grade teachers, Judy Lombard, Dolores Iadoresta and Jean Medick, and the school secretary, Florence Peacock. We are also grateful to the children who participated and to the parents, without whose permission we could not have proceeded.

example, Baddeley, 1978; Kleiman, 1975; Levy, 1977; Liberman, Mattingly, & Turvey, 1972; Tzeng, Hung, & Wang, 1977). In several studies of beginning readers, we and other investigators (Byrne & Shea, 1979; Liberman, Shankweiler, Liberman, Fowler, & Fischer, 1977; Mark, Shankweiler, Liberman, & Fowler, 1977) have found new support for the involvement of phonetic representation in the reading process: The ability to make effective use of phonetic representation appears to be correlated with success at learning to read.

The possibility of an association between children's reading ability and their use of phonetic representation was first explored by Liberman, Shankweiler and their colleagues (Liberman et al., 1977) who assessed the role of phonetic representation in memory for letter strings. Using a modification of Conrad's (1964) procedure, they asked good and poor readers in the second grade to recall a string of consonants in which the letter names either rhymed or did not. In both the rhyming and nonrhyming conditions, good readers recalled more items than poor readers. However, the good readers, like Conrad's adult subjects, were greatly penalized by rhyme, whereas the poor readers performed at about the same level on both rhyming and nonrhyming strings. A subsequent experiment (Shankweiler & Liberman, 1976) showed that the same pattern of recall performance occurred whether the items were presented by ear or by eye. The interaction of reading ability and the effect of phonetic confusability has also been demonstrated in the case of recognition memory for isolated words, where good readers show evidence of greater reliance on the use of phonetic representation as a means of remembering words presented in either written (Mark et al., 1977) or spoken (Byrne & Shea, 1979) form. From all of these findings it would seem that underlying the defective performance of poor readers is a problem that extends beyond the act of recoding from print to speech, involving a more general deficit in the use of phonetic coding in working memory.

Consonant strings and isolated words, however, are far removed from actual text. It remains to be determined whether good and poor readers' recall of more natural linguistic stimuli will be affected by the same variables that affect the recall of letters and words. Accordingly, in this investigation we have extended our study of the effect of phonetic confusability to the more ecologically valid situation of recall of sentences.

Previous findings in the research literature lead us to expect that poor readers' recall of both sentences (Mattis, French, & Rapin, 1975; Perfetti & Goldman, 1976; Pike, Note 2; Weinstein & Rabinovitch, 1971; Wiig & Roach, 1975) and word strings (Bauer, 1977; Katz & Deutsch, 1964) will be inferior to that of good readers. We might suppose from the results obtained in the case of letter strings (Liberman et al., 1977), that the introduction of phonetically confusable words into the sentence or word string to be recalled would differentially affect children who differ in reading ability. We therefore assessed children's ability to recall sentences which vary not only along the traditional dimensions of syntax and meaning (as in Miller & Isard, 1964), but also in the presence of phonetically confusable words. Our materials included seven different syntactic constructions, each of which is presented in four versions: a meaningful version in which none of the words rhyme, a meaningful version in which the majority of words rhyme, a meaningless version in which the words do not rhyme, and a meaningless version in which most words again rhyme. The recall of word strings is examined in an analogous fashion, with items containing five words selected from the meaningless versions of the test sentences. In half of these the words do not rhyme; in half, they do rhyme.

METHOD

Subjects

The subjects were second grade children from a public school in suburban Connecticut. An initial subject pool of 15 good and 15 poor readers was obtained by means of teacher recommendations and scores on the word recognition subtest of the Comprehensive Test of Basic Skills (1974), which had been administered at the end of the first grade. The reading ability of subjects selected in this way was assessed by administration of the Word Attack and Word Identification subtests of the Woodcock Reading Mastery Tests (Woodcock, 1973). The mean sum of raw scores on these subtests was 54.2 for good readers, as compared to 133.9 for poor readers, $\underline{t}(28)=18.19$, $\underline{p}<.001$. There was no overlap between scores of the two groups. The subjects had IQ scores ranging between 90 and 135 on the Slosson Intelligence Test (Slosson, 1963). The mean IQ score for good readers (114.7) was marginally superior to that of poor readers (107.6), $\underline{t}(28)=1.6$, $\underline{p}<.06$. The two groups were not significantly different in mean age: 96.3 months for the good readers, 97.1 for the poor readers. All children had been screened by the school system and found to be free from speech or hearing disorders.

Materials

Sentences. Items for the sentence repetition task were permutations of seven 13-word sentences of English. These seven base sentences were chosen to represent a variety of English constructions with complexity varied along a number of syntactic dimensions. The adoption of 13 words as the sentence length was motivated by a desire to prevent good readers from achieving ceiling performance, since ceiling performance confounds interpretation of many previous studies of the sentence recall of good and poor readers. Before designing the sentence repetition materials, we conducted a pilot study of the effect of sentence length on the sentence recall of eight average readers in the second grade classrooms from which the subjects were drawn. Results indicated that average readers begin to make errors as the length of a meaningful, phonetically nonconfusable sentence approaches 13 words.

Each sentence was presented in four versions, which were constructed by substitutions among content words with position and choice of function words held constant. Thus syntactic structure was the same across the four versions of each base, while manipulations of content words permitted orthogonal variation of sentence meaning and phonetic confusability. Versions were either meaningful, phonetically nonconfusable; meaningful, phonetically confusable; meaningless, phonetically nonconfusable; or meaningless, phonetically confusable. A representative example of a base sentence and its four versions is given in Table 1.

Table 1

BASE SENTENCE:
/NOUN/'s /ADJ/ /NOUN/ /VERB/ (past tense) at the /NOUN/ that /VERB/ (past tense) on the /ADJ/ /NOUN/.

Versions:
Meaningful, phonetically nonconfusable:

Peg's brown dog bit at the bone that fell on the clear floor.

Meaningful, phonetically confusable:

Pat's bad cat bat at the rat that sat on the flat mat.

Meaningless, phonetically nonconfusable:

Bob's fried cap laughed at the chair that stood on the smart glass.

Meaningless, phonetically confusable:

Kay's gray hay stayed at the clay that lay on the gay day.

All versions of each base sentence were matched with respect to word frequency (Thorndike & Lorge, 1944) and the number of syllables contained in each word. The meaningless versions differed from meaningful ones with respect to whether choice of nouns, verbs and adjectives adhered to semantic restrictions. Meaningful versions were created in accordance with these restrictions; meaningless versions were created by violating them. The phonetically nonconfusable and phonetically confusable versions differed with respect to the presence of rhyming items. Phonetically nonconfusable versions contained no rhyming words, phonetically confusable versions contained from seven to nine rhyming words. The number of rhyming words and their position were held constant across the two phonetically confusable versions of each base.

Word strings. The word strings consisted of words obtained from the meaningless, phonetically nonconfusable version and the meaningless, phonetically confusable version of the sentences used in the sentence repetition task. For each string, a set of five words was chosen from among the one-syllable content words of one version. (In the case of phonetically confusable versions, choice was limited to rhyming words.) Each set of five words was then rearranged to form an agrammatic sequence, a manipulation which resulted in a final set of 14 five-item agrammatic word strings, seven of which contained words that rhymed, and seven of which were made up entirely of nonrhyming words.

Procedure

Both the sentence repetition and word string repetition tasks were conducted within a single 20 minute session, with all subjects receiving the sentence repetition materials first. Transcriptions of each subject's responses were made during the experimental session by the examiner and later checked against a tape recording of the child's responses.

The test session was preceded by a training procedure designed to assure that the child understood the task. The examiner explained:

> "I want you to listen to a sentence and then to try to repeat it as best you can. I'll say each sentence twice---the first time I would like you just to listen, but after the second time you hear it, you should try to repeat it. Some of the sentences may seem strange. Sometimes you may find it hard to remember all of the words. It's important for you to say as many of the words as you can remember, even if you have to guess or skip over parts you don't remember. Let's try a few, just for practice. I'll read each sentence twice. After the second time I read it, you try to repeat it. Ready?"

Following the instructions, the child was presented with a set of four practice items: two 13-word meaningful, phonetically nonconfusable sentences and two 13-word meaningless, phonetically nonconfusable ones. The experimenter read each sentence twice, after which the child was asked to repeat the sentence. If the child made no attempt to respond, the sentence was read a third time; children who hesitated over a word were encouraged to guess or to skip over that word. On completion of the four practice items, the child was advised:

> "Now I am going to use the tape recorder to play some more sentences for you. This time, a man will say the sentence. He'll say each sentence twice, just as I did. Remember, try to say the sentence after you hear it the second time. Say as many words as you can."

At this time, a pre-recorded series of the test sentences was played to the child. The series included four versions of each of the seven base sentences, arranged in a fixed random order. Each sentence was repeated twice by a male, native American speaker of English, who attempted to hold prosody constant across the four versions of each base sentence. During actual testing, there was no prompting for responses, nor were unrecalled sentences repeated a third time.

The sentence repetition and word string repetition tasks were separated by a brief rest period. During this break, the examiner explained:

> "Now I am going to play five words for you, one at a time. Listen to them carefully, because you will hear them only one time. After you have heard all five words, try to say them back in the same order. Remember, say as many words as you can, and guess if you have to."

The examiner then played a pre-recorded set of the 14 five-item word strings. Like the sentences, they were presented in a fixed random order, and were spoken by the same male speaker. However, unlike the sentences, each string was read only once. Words within the string were read at the rate of one per sec with prosody held neutral.

Scoring Procedure

Sentences. The error scores were the sum of omissions, substitutions and reversals made on each version of each base sentence. All versions were scored in the following manner: A score of 0 was given for correct repetition with no errors. One point was given for each word recalled in the improper sequence (relative to the preceding word), for each substitution, and for each intrusion. Words that followed substitutions or intrusions were scored relative to the immediately preceding word that had been a member of the original sentence. A score of 13 was given when a subject failed to repeat any of the words of the sentence.

Word strings. For word strings, as for sentences, the error score was the sum of omissions, substitutions and reversals. To minimize the effects of guessing, only the first five words produced during recall were counted. A score of 0 was given if all items were recalled in proper order. One point was given for each word recalled in the improper order, and for each substitution and intrusion. Words preceded by a substitution were scored relative to the immediately preceding member of the sequence. A score of five was given if the subject failed to recall any of the items.

RESULTS

This experiment was conducted to determine whether the verbal memory of good and poor readers would be differentially affected by systematic variations in phonetic confusability of the material to be recalled. For this purpose, memory for sentences and for agrammatic word strings was examined separately. The effects of systematic variations in meaningfulness were also examined in the case of sentence memory, as was the effect of variations in syntactic structure.

Sentence Repetition

In considering sentence memory, we needed first to ascertain that our good and poor readers could be differentiated by their overall performance on our materials. To this end, error score data on all the sentence versions were subjected to an analysis of covariance with IQ controlled. It was found, as expected, that good readers made fewer errors overall than poor readers. The mean error score for good readers was 4.7, as compared to 5.3 for poor readers, $F_{IQ}(1,27)=7.6$, $p<.01$. Another prior requirement was to determine whether the stimulus variations we had introduced had any differential effects on the performances of the two groups. Each of the sentences had been presented in four versions which varied orthogonally in phonetic confusability and meaningfulness. As can be seen in Figure 1, in which mean error scores on each version type are separately plotted, children made more errors on rhyming versions than on nonrhyming versions, $F_{IQ}(1,28)=124.5$, $p<.001$. Meaningfulness also produced a significant effect, $F_{IQ}(1,28)=172.6$, $p<.001$.

Having established that the good and poor readers did indeed differ in sentence memory, and that they were both affected by the stimulus variations, we turned next to the central focus of this study, which is the interaction between these stimulus variables and reading ability. We found that good readers were affected by phonetic confusability to a markedly greater extent than poor readers, $F_{IQ}(1,27)=90.9$, $p<.001$. No such interaction was obtained for the variable of meaningfulness. Supplementary analysis by t-test permitted us to further assess mean differences between the two groups on each of the four sentence versions. Here, as in the overall analysis, good readers made significantly fewer errors than poor readers on the phonetically non-confusable versions, both meaningful $t(12)=4.2$, $p<.001$; and meaningless $t(12)=5.1$, $p<.005$. In contrast, when the items were phonetically confusable, the performance of the good readers actually dropped to the level of that of the poor readers. Thus, in recall of the rhyming versions, both meaningful and meaningless, the performance of the good and poor readers did not differ significantly, as is depicted in Figure 1.

An analysis was conducted to examine the consistency of the effects of phonetic confusability and meaningfulness across the seven base sentences. In order to ascertain whether some few sentences contributed disproportionately to the main effects revealed by the analysis of covariance, we compared performance among the four versions of each base. More errors were made on rhyming versions than on nonrhyming ones for six of the seven base sentences, and more errors were made on meaningless versions than on meaningful ones for all seven base sentences. Analysis of variance reveals a significant interaction of phonetic confusability and type of base sentence, $F(1,168)=5.9$, $p<.001$, and a significant interaction of meaningfulness and type of base sentence, $F(1,168)=8.2$, $p<.001$. However, there is no three-way interaction of reading ability, phonetic confusability and type of base sentence, or of reading ability, meaningfulness and type of base sentence.

An additional analysis was carried out to treat base sentence as a random variable nested within phonetic confusability and meaningfulness (see Clark, 1973). A significant interaction of reading ability and phonetic confusability was upheld, $minF'_{IQ}(1,31)=4.3$, $p<.05$; but there was no significant interaction of reading ability and meaningfulness.

We turned finally to compare performance across the seven base sentences, a comparison which is not central to our purposes, but is nevertheless permitted by our design. Since the base sentences were chosen to vary along a number of syntactic dimensions, it was expected that error rates in recalling them would differ. This expectation was confirmed, $F(1,168)=29.3$, $p<.001$. There was, however, no significant interaction of reading ability and the effects of base sentence, showing that good and poor readers in our sample were comparably affected by the syntactic variations. A comparison of the distribution of errors made by good and poor readers on each of the four versions of each sentence provides further evidence that the two groups reacted similarly to variations in syntactic structure. The frequency of errors as a function of the position of words in the sentence was significantly correlated for the two groups in most versions, $r(13)>.46$, $p<.05$ for 26 of

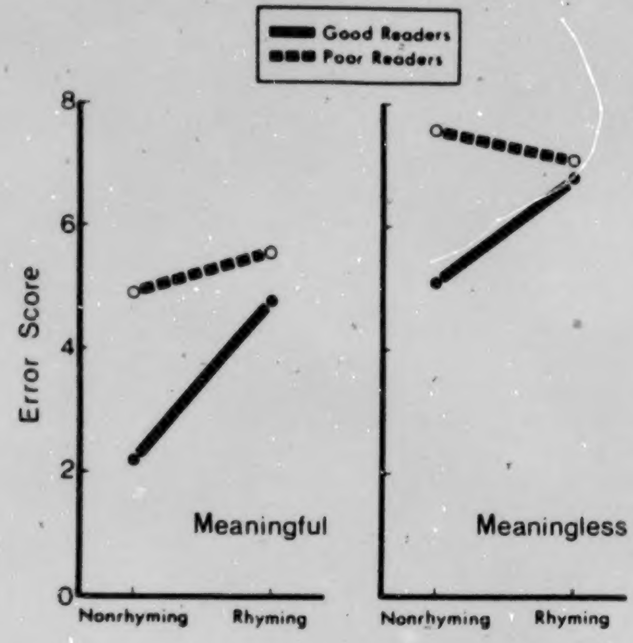


Figure 1. Good and poor readers' mean performance on meaningful and meaningless sentence versions, in nonrhyming and rhyming conditions.

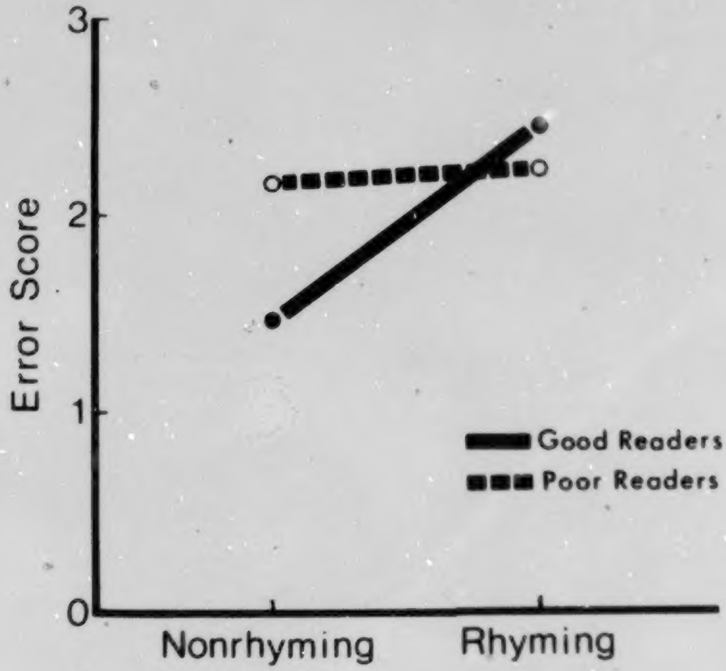


Figure 2. Good and poor readers' mean performance on word strings in nonrhyming and rhyming conditions.

the 28 versions; $\underline{r}(13)>.68$, $\underline{p}<.005$ for 21 of them. Thus the errors of good and poor readers are similarly distributed, differing only in frequency of occurrence.

Word strings

As was the case with the sentence repetition data, error scores on word string recall were subjected to an analysis of covariance with IQ controlled. Mean scores for good and poor readers are plotted in Figure 2. It may be seen that an overall difference in error score was again found for good and poor readers, $\underline{F}_{IQ}(1,27)=4.50$, $\underline{p}<.05$. Also apparent once more is a significant effect of phonetic confusability, $\underline{F}(1,28)=12.8$, $\underline{p}<.002$: children made more errors in recall of rhyming strings. The crucial interaction of reading ability and the effect of phonetic confusability is again strongly manifest, $\underline{F}_{IQ}(1,27)=9.5$, $\underline{p}<.002$. As illustrated in Figure 2, the performance of good readers was markedly impaired by phonetic confusability while that of the poor readers was not.

A test was made of the generality of these findings by an analysis of variance with word string treated as a random variable. Here, as in the preceding analysis of covariance, the interaction of reading ability and phonetic confusability is significant: $\underline{minF'}_{IQ}(1,14)=5.71$, $\underline{p}<.05$.

DISCUSSION

As we noted in the introduction, a number of studies in the research literature report that unskilled readers tend to perform more poorly than skilled readers in short-term recall of letter strings, word strings, and sentences. In studies of letter-string recall (Liberman et al., 1977; Shankweiler & Liberman, 1976), demonstrations of the greater vulnerability of good readers to the penal effects of phonetic confusability suggest that these children place greater reliance on phonetic coding as a short-term memory strategy. Correspondingly, the demonstration that recall in poor readers was little affected by the phonetic characteristics of the items suggests that they are making ineffective use of phonetic coding in working memory. Our aim in the present study was to test the generality of this interpretation by asking whether phonetic confusability will also differentially affect good and poor readers' recall not only of alphabetic strings but also of sentences and word strings.

To this end, good and poor readers in a second grade sample were asked to repeat specially designed sentences and agrammatic word strings. Consistent with previous reports, good readers were better than poor readers when the material to be recalled, whether sentences or word strings, contained no phonetically confusable words. In contrast, the performance of good readers fell to the level of poor readers when phonetically confusable words were present. Although some studies have found that poor readers are more adversely affected than good readers by manipulations that destroy meaningfulness (Pike, Note 1, Note 2; Wiig & Roach, 1975), our systematic variations of meaningfulness and syntactic structure did not differentially affect the two reading groups. The primary distinction, once again, was that good readers were severely impaired by the introduction of phonetic confusability and the poor readers were not.

These findings confirm the results of Liberman et al. (1977) and extend them to the more natural task of sentence recall. Since the same pattern of interaction with phonetic confusability has been found for three different classes of items--letters, words and sentences--a common etiology is implicated. We follow Liberman et al. (1977) in suggesting that the poor readers' substandard recall of verbal material may be caused by failure to make effective use of phonetic coding in working memory.

We have viewed these and other findings of correlation between effective use of phonetic coding and success at learning to read as further indication of the ubiquitous involvement of speech coding in the reading process. It could be supposed, however, that ineffective phonetic coding is a by-product rather than a determinant of reading difficulty. This question might be laid to rest if it could be shown that deficient use of phonetic coding in preschool children is predictive of reading failure, both in English and in languages that manifest quite different morphologies and writing systems. We are in the process of gathering data pertinent to this issue.

Other investigators have commented on the association between reading difficulty and deficient verbal short-term memory (see, for example, Perfetti & Goldman, 1976; Perfetti & Lesgold, 1979; Vellutino, Steger, DeSetto, & Phillips, 1975; Vellutino, Steger, Kaman, & DeSetto, 1975). Moreover, we are not alone in supposing that these deficiencies apply to perception of language by ear as well as by eye. Our supposition that a number of memory related problems may be seen as manifestations of deficient phonetic coding (Shankweiler, Liberman, Mark, Fowler, & Fischer, 1979) is consistent with the views of Perfetti and his colleagues. It is appropriate at this point to consider what precisely might be the basis of the poor reader's limitations in use of phonetic representation. In a recent paper (Shankweiler et al., 1979), we raise the question whether the deficits may extend beyond the memorial aspects of language, involving perhaps the level of perceptual encoding. If so, then sufficiently stringent tests of speech perception might be expected to distinguish good and poor readers of the sort studied here. We are currently investigating this possibility, bearing in mind the hypothesis of Perfetti and Lesgold (1979) that the short-term memory differences between good and poor readers may largely derive from slower encoding on the part of poor readers.

At the same time that we are led to consider the basis of the poor reader's ineffective use of phonetic coding, we are also led to speculate as to its broader implications. Here we are guided by the assumption that a major function of phonetic coding in both written and spoken language is to facilitate interpretation of stretches of discourse longer than the word. If poor readers do, in fact, fail to make effective use of phonetic coding, then they may have difficulty comprehending some kinds of sentences in situations in which working memory is stressed.

We conclude by suggesting two ways in which working memory might be stressed in sentence processing. First, it may be stressed when recovery of syntactic structure requires retention of many component words of a sentence.

Such could be the case in center-embedded sentences and sentences involving extensive movement or deletion (cf. Frazier & Fodor, 1978; Kimball, 1975, for a discussion of sentence parsers). Accordingly, these might pose more difficulty for poor readers than for good readers. Second, even when syntactic structure is relatively simple, working memory may be stressed if word order is in some way crucial. The importance of word order in this sense has been discussed by Baddeley (1978) and is exemplified in the Token Test of DeRenzi and Vignolo (1962). We suspect that Token Test instructions such as "Touch the large, red triangle with the small, green square," might differentiate between good and poor readers, and we intend to pursue this possibility.

REFERENCE NOTES

1. Pike, R. *Linguistic development as a limiting factor in learning to read*. Paper presented at the First Annual Boston University Conference on Language Development, Boston, Mass., 1976.
2. Pike, R. *Memory for words and reading ability*. Paper presented at the Biennial Meeting of the Society for Research in Child Development, New Orleans, La., 1977.

REFERENCES

Baddeley, A. D. The trouble with levels: A reexamination of Craik and Lockhart's framework for memory research. *Psychological Review*, 1978, *85*, 139-152.

Bauer, R. Memory processes in children with learning disabilities. *Journal of Experimental Child Psychology*, 1977, *24*, 415-430.

Byrne, B., & Shea, P. Semantic and phonetic memory in beginning readers. *Memory & Cognition*, 1979, *7*, 333-338.

Clark, H. H. The language-as-fixed-effect fallacy: A critique of language statistics in psychological research. *Journal of Verbal Learning and Verbal Behavior*, 1973, *12*, 335-359.

Comprehensive Test of Basic Skills. Monterey, Ca.: CTB/McGraw Hill, 1974.

Conrad, R. Acoustic confusions in immediate memory. *British Journal of Psychology*, 1964, *55*, 75-84.

DeRenzi, E., & Vignolo, L. A. The Token test: A sensitive test to detect receptive disturbances in aphasia. *Brain*, 1962, *85*, 665-678.

Frazier, L., & Fodor, J. D. The sausage machine: A new two-stage parsing model. *Cognition*, 1978, *6*, 291-325.

Katz, P. A., & Deutsch, M. Modality of stimulus presentation in serial learning for retarded and normal readers. *Perceptual and Motor Skills*, 1964, *19*, 627-633.

Kleiman, G. M. Speech recoding in reading. *Journal of Verbal Learning and Verbal Behavior*, 1975, *14*, 323-339.

Kimball, J. Predictive analysis and over-the-top parsing. In J. Kimball (Ed.), *Syntax and semantics* (Vol. 4). New York: Academic Press, 1975.

Levy, B. A. Reading: Speech and meaning processes. *Journal of Verbal Learning and Verbal Behavior*, 1977, *16*, 623-638.

Liberman, A. M., Mattingly, I. G., & Turvey, M. Language codes and memory codes. In A. W. Melton & E. Martin (Eds.), *Coding processes and human memory*. Washington, D.C.: Winston and Sons, 1972.

Liberman, I. Y., Shankweiler, D., Liberman, A. M., Fowler, C., & Fischer,

F. W. Phonetic segmentation and recoding in the beginning reader. In A. S. Reber & D. Scarborough (Eds.), Toward a psychology of reading: The proceedings of the CUNY Conference. Hillsdale, N.J.: Erlbaum, 1977.

Mark, L. S., Shankweiler, D., Liberman, I. Y., & Fowler, C. A. Phonetic recoding and reading difficulty in beginning readers. Memory & Cognition, 1977, 5, 623-629.

Mattis, S., French, J. H., & Rapin, I. Dyslexia in children and young adults: Three independent neuropsychological syndromes. Developmental Medicine and Child Neurology, 1975, 17, 150-163.

Miller, G. A., & Isard, S. Free-recall of self-embedded English sentences. Information and Control, 1964, 7, 292-303.

Perfetti, C. A., & Goldman, S. Discourse memory and reading comprehension skill. Journal of Verbal Learning and Verbal Behavior, 1976, 14, 33-42.

Perfetti, C. A., & Lesgold, A. M. Discourse comprehension and sources of individual differences. In M. Just & P. Carpenter (Eds.), Cognitive processes in comprehension. Hillsdale, N.J.: Erlbaum, 1979.

Shankweiler, D., & Liberman I. Y. Exploring the relations between reading and speech. In R. M. Knights & D. K. Bakker (Eds.), Neuropsychology of learning disorders: Theoretical approaches. Baltimore: University Park Press, 1976.

Shankweiler, D., Liberman, I. Y., Mark, L. S., Fowler, C. A., & Fischer, F. W. The speech code and learning to read. Journal of Experimental Psychology: Human Learning and Memory, 1979, 5, 531-545.

Slosson, R. L. Slosson Intelligence Test for children and adults. East Aurora, N.Y.: Slosson Educational Publications, 1963.

Thorndike, E. L., & Lorge, I. The teacher's word book of 30,000 words. New York: Teachers' College Press, 1944.

Tzeng, O. J. L., Hung, D. L., & Wang, W. S-Y. Speech recoding in reading Chinese characters. Journal of Experimental Psychology: Human Learning and Memory, 1977, 3, 621-630.

Vellutino, F. R., Steger, J. A., DeSetto, L., & Phillips, F. Immediate and delayed recognition of visual stimuli in poor and normal readers. Journal of Experimental Child Psychology, 1975, 19, 223.

Vellutino, F. R., Steger, J. A., Kaman, M., & DeSetto, L. Visual form perception in deficient and normal readers as a function of age and orthographic/linguistic familiarity. Cortex, 1975, 11, 22-38.

Weinstein, R., & Rabinovitch, M. S. Sentence structure and retention in good and poor readers. Journal of Educational Psychology, 1971, 62, 25-30.

Wiig, E. H., & Roach, M. A. Immediate recall of semantically varied sentences by learning disabled readers. Perception and Motor Skills, 1975, 40, 119-125.

Woodcock, R. W. Woodcock Reading Mastery Tests. Circle Pines, Mn.: American Guidance Services, Inc., 1973.

EFFECTS OF VOCALIC FORMANT TRANSITIONS AND VOWEL QUALITY ON THE ENGLISH [s]-[š] BOUNDARY*

D. H. Whalen+

Abstract. In two experiments, the effects of the vocalic portion of fricative-vowel syllables on the perception of alveolar and palatal fricatives were examined. The fricatives were synthesized to represent a continuum from [s] to [š]; the vowels ranged from [u] to [i] through [ï] and [ü]. The vocalic formant transitions were of two types, those appropriate to [s] and those to [š]. All stimuli were presented in forced-choice labelling tests. The boundary between [s] and [š] for English-speaking listeners varied as a function both of transitions and of vowel. The effect of the transitions on the s/š boundary was clear and straightforward: An ambiguous noise was heard more often as [s] before [s] transitions, and as [š] before [š] transitions. The quality of the vowel clearly had an effect, but this effect was open to more than one interpretation. The responses of listeners who were unfamiliar with languages that use [ü] and/or [ï] distinctively were not significantly different from those of listeners who were familiar with such languages.

INTRODUCTION

Although an obvious cue for the place of articulation of fricatives lies in the spectrum of the aperiodic noise, other cues are identifiable. One such cue is provided by the formant transitions of the following vocalic segment, whose effect was demonstrated in a preliminary way by Harris (1958). In that experiment, natural speech syllables composed of English fricatives followed by various vowels were copied on magnetic tape and cut into fricative and vocalic segments which were then recombined so that each noise occurred with each vocalic portion (including the transitions appropriate for each of the places of articulation). Transitions were found to determine the perceived place of articulation for the dental and labial fricatives [f,θ,v,ð], but the alveolar and palatal noises [s,š,z,ž] seemed strong enough cues to outweigh any cues contained in the transitions. This result has been replicated

---

*Part of this research was presented at the December 1978 Meeting of the Linguistic Society of America, Boston.

+Also Yale University.

Acknowledgment: This research was supported by NICHD Grant HD01994 and BRS Grant RR05596 to Haskins Laboratories. I thank Alvin Liberman and Bruno Repp for their helpful suggestions and criticism. Guy Carden, Terry Halwes, Virginia Mann, and P. J. Price provided useful comments on earlier versions of this paper.

[HASKINS LABORATORIES: Status Report on Speech Research SR-59/60 (1979)]

(LaRiviere, Winitz, & Herriman, 1975); however, McCasland (1978) recently reported that, for fricatives in intervocalic position, transition cues may occasionally override the cues provided by the [s] and the [š] noise. Since the natural noise, at least in utterance-initial position, is such a powerful cue, tape-splicing experiments are inherently insensitive to possible effects on perception due to the alveolar and palatal transitions. Such effects were in fact found by Delattre, Liberman, and Cooper (1964) for the voiced fricatives [z] and [ž] when, in synthetic speech, a neutral noise was used before varying transitions. The present experiments were designed in part to provide a more sensitive method so that, if the transitions affect the perception of the voiceless alveolar and palatal fricatives as well, a rough measure of the magnitude of the effect could be found.

Another goal of the present study was to see if the quality of a following vowel might not also affect the perception of a preceding fricative. Such differences in perception (reflected in the Japanese [s]-[š] boundaries measured on a synthetic noise continuum when the fricatives were placed before different vowels) were found in experiments by Kunisaki and Fujisaki (1977). (The nature of a preceding vowel was found to have a much smaller effect on the boundary.) Their results for the vowels [a e o u] suggested that the important feature of the vowel was rounding. An ambiguous noise was more likely to be heard as [š] when it preceded one of the unrounded vowels [a,e] than when it preceded one of the rounded ones [u,o]. (In Japanese, [s] does not occur before [i].) Since, in articulation, lip rounding in anticipation of the rounded vowels would lower the frequency of a preceding fricative noise, this result suggests "that the influence of context in speech production is corrected in speech perception" (Kunisaki & Fujisaki, 1977, p. 91). The present experiments were intended to replicate these findings with a somewhat different set of vowels.

The final purpose of the present experiments was to determine whether subjects who do not use certain vowels distinctively would still show a unique effect of each such vowel on the perception of a preceding fricative. Since the four vowels chosen for the first experiment included two that do not occur phonemically in English, it was possible to test subjects whose speech did not contain the foreign vowels [ü,ï] and who had not studied languages that do contain them.

EXPERIMENT 1

To determine the boundary between [s] and [š], a fricative noise continuum, with stimuli systematically varied in the frequency of two formants, was used. Thus, for the requisite control, synthetic noises were needed. However, the vocalic segments needed to vary only by vowel quality and by transitions. Since natural speech controls the transitions automatically in the [s] and [š] environments, and since vowel quality can be controlled to some extent by training, natural-speech vocalic segments were used in Experiment 1.

Method

A linguist familiar with languages using all four of the vowels to be used pronounced eight syllables: each of the vowels [i, u, ï, ü] preceded by [s], and by [š]. These utterances were recorded on magnetic tape and were then digitized using the Haskins Laboratories pulse code modulation (PCM) system. The fricative noises were cut off with the aid of magnified waveform displays, and the resulting eight vocalic segments were then combined with the noises of a ten-member fricative continuum produced on the OVEIIIc synthesizer at Haskins Laboratories. The synthetic fricatives were 200 msec in length and differed in the frequencies of the two fricative formants, which increased in approximately equal steps from the lower, more [š]-like noises to the higher, more [s]-like noises (see Table 1). The bandwidths varied from approximately 100 Hz for the lowest frequencies to approximately 800 Hz for the highest.

Table 1

Fricative Formants for Experiment 1, in Hz

	Stimulus Number									
	1	2	3	4	5	6	7	8	9	10
F2	3020	3488	4030	4655	5226	5699	6397	6778	7391	8061
F1	1008	1307	1646	2074	2328	2613	3019	3293	3591	4032

When the results for the test were tabulated, it became clear that the subjects were not reliably perceiving the first five stimuli of the continuum either as [s] or [š]. While there was a bias in the response toward whichever fricative was appropriate to the transition that was present, many of the data points were at or near chance. Presumably, this occurred because the frequencies of these noises were too low to give an acceptable [š]. Therefore, the results from these five stimuli will be omitted for clarity of presentation.

The ten fricatives were combined with the eight vocalic segments (four vowels, each with [s] or [š] transitions). The resulting 80 stimuli were presented in randomized order, 6 repetitions per session, 2 sessions per subject. The subjects were five Yale students who were native speakers of English and who were unfamiliar with languages containing the foreign vowels [ï] and [ü]. The subjects were asked only to identify the fricative as "s" or "sh". Stimuli were presented binaurally over Telephonics TDH-39 earphones at the rate of one every three seconds.

In order to check that the two foreign vowels were in fact identified as either [i] or [u] and not some other English vowel, a vowel identification test was conducted at the end of the last session. The eight experimental utterances, including the original frication, were presented together with ten more syllables, [s] and [š] before [e, ö, o, ë, a], all produced by the same talker. Five repetitions of each stimulus were presented in randomized order. In order to avoid using special symbols that would necessitate training the subjects, ten English words were chosen for the response sheets. The words were: see/she, say/Shea, saw/Shaw so/show, sue/shoe. The results of the identification test were quite consistent. [i] and [u] were always heard correctly, and [ü] was heard as [i] and [ï] as [u] over 95% of the time.

Results

In Figure 1, the average percentage of "sh" responses is plotted against the frequency of the first fricative formant. For each of the four vowels, the responses for the [s] transition stimuli are presented separately from those for [š] transition stimuli. One can see that in each case, there were far fewer "sh" responses when the transition was appropriate to [s] than when it was appropriate to [š]. The most extreme case is that of [ï] with [s] transitions, for which "sh" responses never got above 28%. It is unclear why this particular vowel/transition combination should be such an overriding cue. The shapes of the response functions in Figure 1 also suggest that stimulus 6 (with a first-formant frequency of 2613 Hz) was actually the best [š] noise, since it was most effective in competing against the conflicting [s] transitions. (Stimuli 1-5 all had lower percentages of "sh" responses.)

Although it is clear from Figure 1 that the four vowels differed in their effects on fricative perception, these effects are difficult to compare. In Figure 2, the same data have been replotted, but with the results for [s] transitions in Figure 2a, and those for [š] transitions in Figure 2b. The effect of the vowel is now apparent in the varying points at which the response is evenly divided between "s" and "sh" (which point will be defined as the s/š boundary). With [š] transitions, the effect of the vowel can be as large as 500 Hz, and it would probably have been similarly large for the [s] transitions if the response had approached 100% for all the vowels, or even if "sh" response for [ï] had reached 50% at all. For each transition, the percentage of "sh" responses was much less for [ï] than for the other three vowels. This was surprising since [ï] was expected to be acoustically intermediate between [i] and [u] and to have a correspondingly intermediate effect on the perception of the fricatives.

Analysis of variance was performed on the total number of "sh" responses, since not all vowels had boundaries. The effect of the transitions was significant, $F(1,4) = 23.2$, $p < .01$, as was that of the vowel, $F(3,12) = 17.4$, $p < .001$. In addition, there was a significant interaction between vowel and transition effects, $F(3,12) = 3.7$, $p < .05$, which is difficult to interpret.

EXPERIMENT 2

While the transitions have a clear effect in Experiment 1, the lack of an s/š boundary for [ï] makes comparative measurements of the transitions' effects impossible. Clear cases of "sh" for all vocalic segments would make numerical comparison meaningful. Also, while there was definitely an effect of the vowel on the perception of the fricative, the results are difficult to interpret. Since Experiment 1 used only one token of each vocalic segment, there is a good possibility that the vowel effects reflected token-specific peculiarities. So, in the second experiment, synthetic vocalic segments were used as well as synthetic fricatives; thus vowel quality could be controlled and systematically varied. In order to facilitate comparison of the two experiments, the same high vowel range ([u] to [i]) was used.

Method

Again using the Haskins OVEIIIc synthesizer, a vowel continuum of eight stimuli was produced, with F1 held constant at 250 Hz, and F2 and F3 systematically varied (see Table 2). These vowels were perceived by the experimenter as ranging from [u] to [i], through [ï] and [ü], with two stimuli in each vowel category.

Table 2

Synthetic Vocalic Segments, Experiment 2, in Hz

	Stimulus Number							
	1	2	3	4	5	6	7	8
F3	2000	2104	2197	2295	2396	2502	2594	2709
F2	800	1000	1198	1404	1600	1796	2001	2198

To model the [s] and [š] transitions of the first experiment, two loci for the second formant were postulated (see Delattre et al., 1964). The most divergent loci, which would nonetheless still give integrable transitions (i.e., not give rise to a "click" percept, as determined by the experimenter), were chosen for the experiment. A satisfactory locus for [s] was found at 1500 Hz, and for [š] at 2000 Hz. Transitions did not reach their loci, but rather started at a point which would be 10 msec after the beginning if the transitions lasted 50 msec. In effect, the transitions were 40 msec long, and pointing at, but not reaching, their loci. The first and third formants were left without transitions; this did not affect perception in any immediately obvious way and reduced the number of variables.

For this experiment a different fricative continuum was synthesized, again with ten members, but covering a somewhat different range (see Table 3). Bandwidths varied from approximately 200 Hz for the lowest frequency to approximately 475 for the highest. The sixteen vowels (eight second-formant values, two transitions each) were combined with the fricatives yielding 160 stimuli. These were randomized as before, and presented to subjects for forced choice labelling as "s" or "sh". Three randomizations, each containing two tokens of each stimulus, were presented in each session. The stimulus rate was one every three seconds. Nine Yale undergraduates with varying amounts of linguistic training (see below) were subjects for three sessions each.

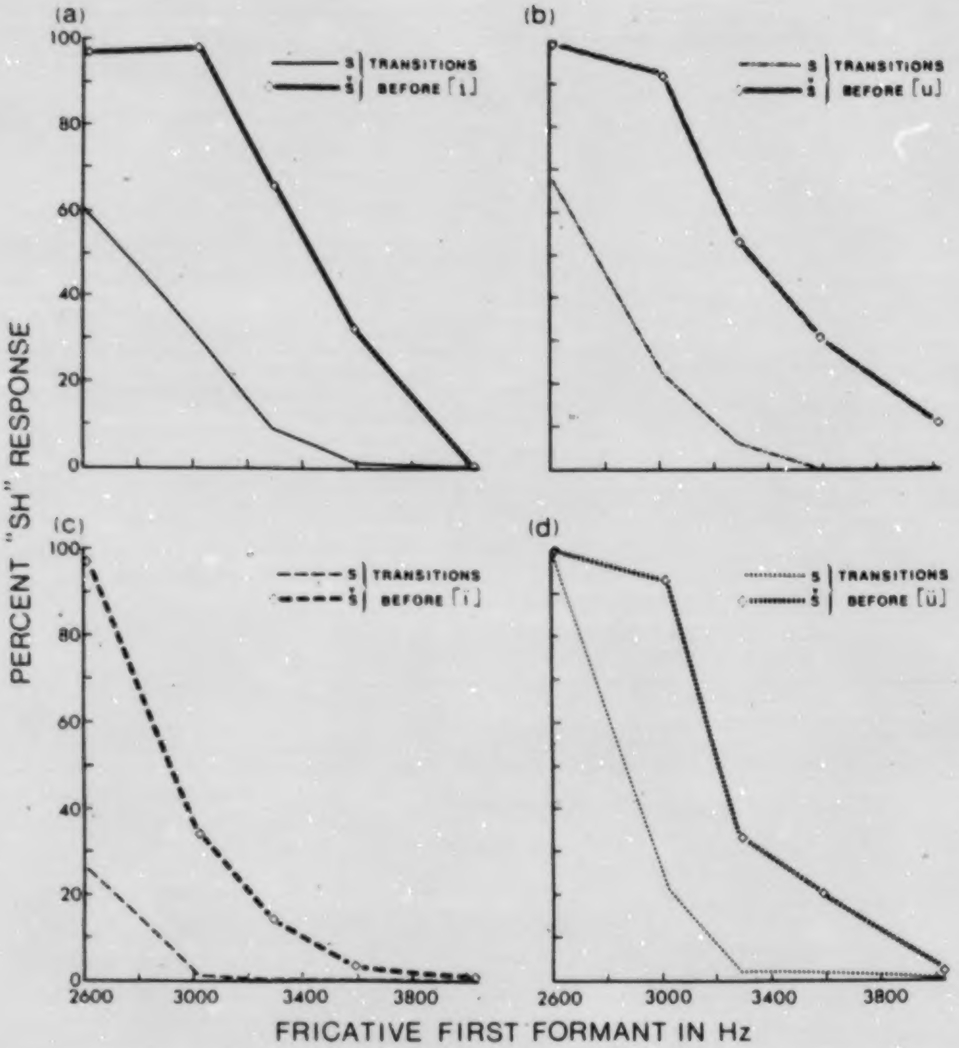


Figure 1. Effect of natural ([s] or [š]) transitions on the perception of synthetic fricatives, plotted separately for each of four vowels. (Experiment 1.)

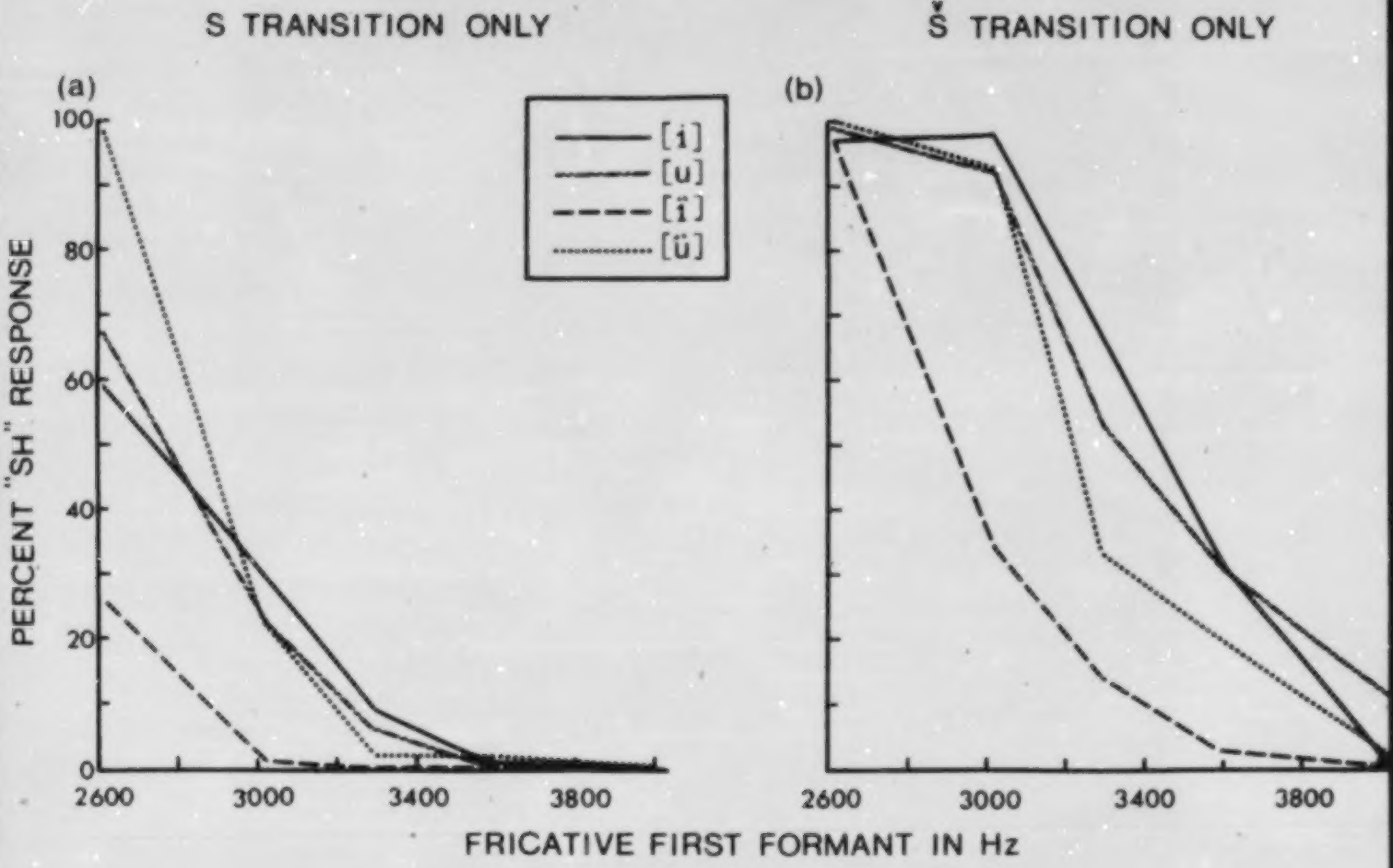


Figure 2. Differential effects of the four vowels on the perception of the fricative. (Experiment 1.)

Table 3

Fricative Formants for Experiment 2, in Hz

	Stimulus Number									
	1	2	3	4	5	6	7	8	9	10
F2	2614	2769	3020	3389	3488	3695	3803	3915	4030	4523
F1	2015	2197	2466	2690	2769	2850	3019	3108	3199	3489

Results

The results from Experiment 2 are presented in Figures 3, 4, and 5. Figure 3 presents each vowel separately, contrasting the [s] and [š] transitions. Here again, as in Experiment 1 (Figure 1), the [s] transition always drew more [s] than [š] responses. Thus the [s] transition response curve is to the left of the [š] transition response curve in Figures 3a-h. Since the s/š boundary is found at the point where the response function crosses the 50% line, this means that the boundary is at a lower noise frequency when [s] transitions follow than when [š] transitions follow. This replicates the results from Experiment 1, although the effect is smaller. However, the effect was highly significant, $F(1,8) = 90.9$, $p < .0001$. Similarly, the effect of the vowel was significant, $F(7,56) = 28.1$, $p < .0001$. There was also an interaction between transition and vowel, $F(7,56) = 5.5$, $p < .0001$. As can be seen in Figure 3 (and even better in Figure 5), the transition effect diminished somewhat at the [i] end of the vowel continuum.

In Figure 4, we can see the effect of the vowel on the boundary. Since there are eight vowels instead of four, the graph is rather difficult to read. Nonetheless, we can see that the change effected by the two most extreme vowels is approximately 200 Hz. While this numerical value is less than that from the natural-speech experiment, it, like that previous result, is of the same order of magnitude as the effect due to the transition.

To make the relative effect of each vowel more apparent, the s/š boundaries were determined for each and plotted in Figure 5. F2 frequency has been chosen as a convenient independent variable, since it was varied systematically. Its use here is not meant to imply that F2 is the relevant parameter for explaining the different effects of the vowels. Both the [s] transition function and the [š] transition function show a general increase in the boundary frequency from the [u]-like sounds to the [i]-like. The [s] transition function shows significant linear and quadratic trends, $F(8,22) = 4.0, 4.9$, $p < .01$, while the [š] transition function has only a significant linear trend, $F(8,22) = 5.0$, $p < .01$.

The subjects were divided into three groups, based on their familiarity with the foreign speech sounds. The most sophisticated group had had at least

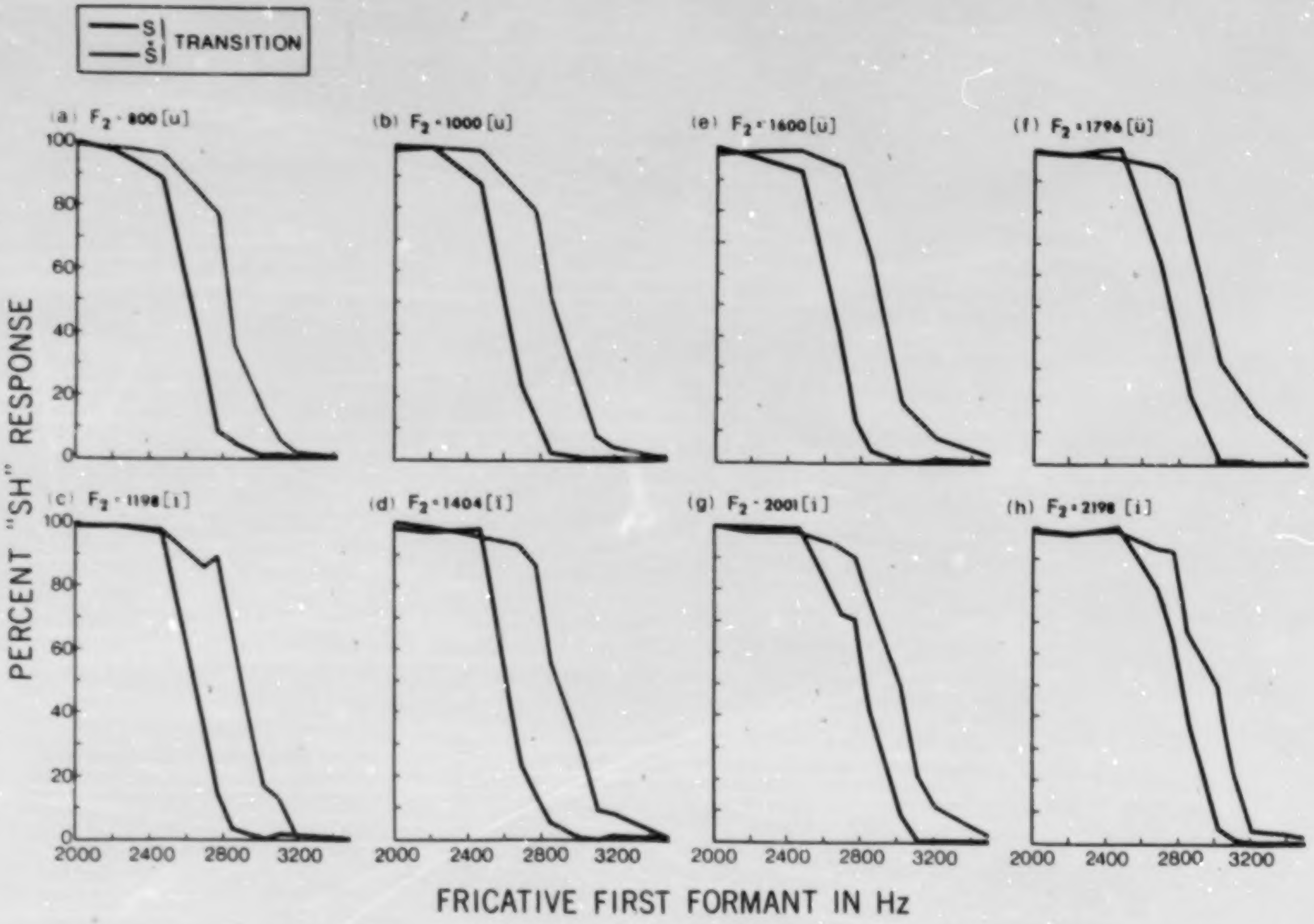


Figure 3. Effect of synthesized [s] and [š] transitions on the perception of the synthetic fricatives. Plotted separately for each of eight vowels, identified both by F2 values and vowel quality as perceived by the experimenter. (Experiment 2.)

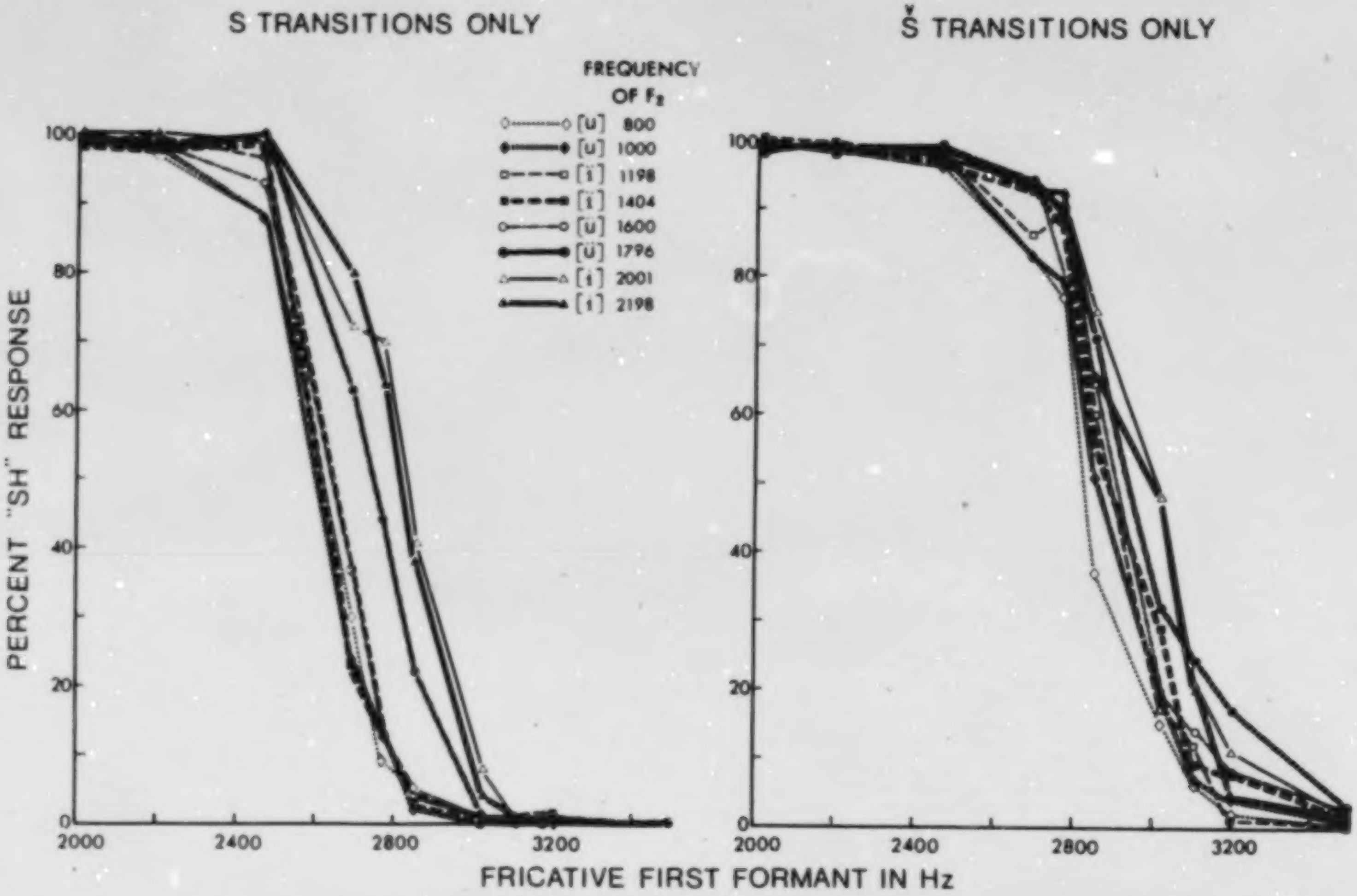


Figure 4. Differential effects of the eight synthetic vowels on the perception of the fricatives. (Experiment 2.)

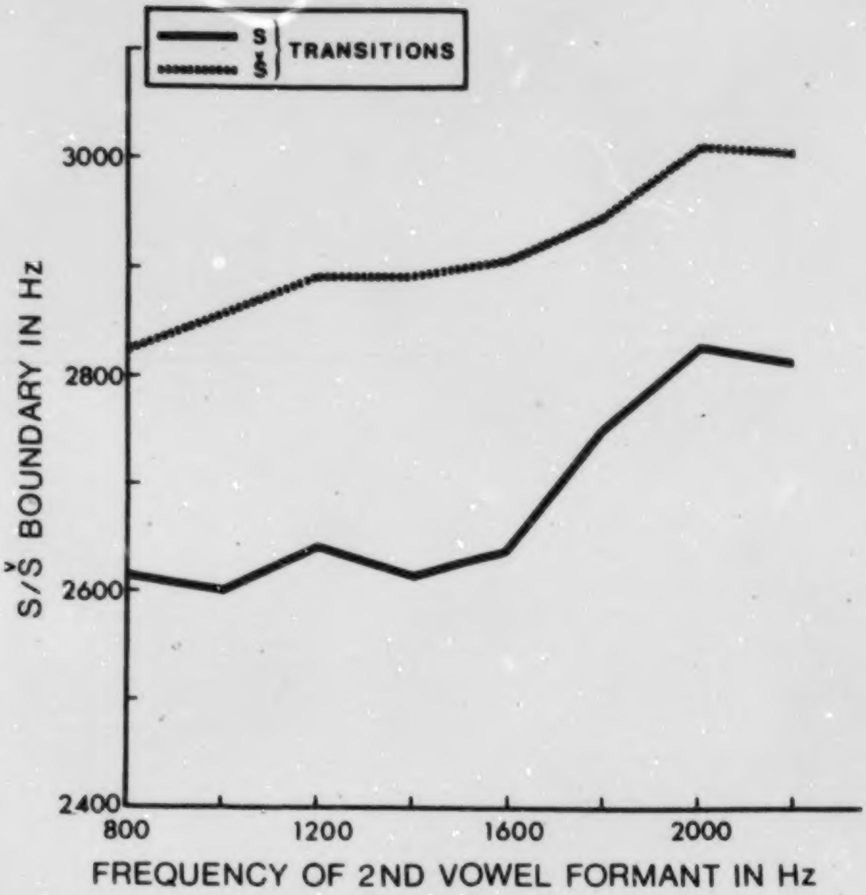


Figure 5. As determined from Figure 4, s/š boundaries plotted (arbitrarily) against F2 of the synthesized vowels. The data for the two transitions are presented separately. (Experiment 2.)

a one-semester course in liguistics, which included articulatory phonetics and transcription of foreign sounds. The three subjects in this group had also studied languages containing [ü]; two were fluent in French. Three other subjects, comprising the second group, had studied languages containing either [ü] or [ï], but had had no specific linguistic training. The last group of three had no linguistic training, nor had they ever studied any languages containing the vowels [ï] or [ü]. Analysis of variance showed no significant difference (at the .05 level) among the responses given by these three groups. For at least these levels of differing linguistic sophistication, familiarity with the foreign vowels did not seem to affect performance.

DISCUSSION

In earlier experiments concerning the effect of vocalic formant transitions on the perception of fricatives, the fricative noise was not systematically varied. Thus it was not possible to determine the boundaries between fricatives, nor the amount of shift in the boundaries due to the transitions. In the present two experiments, we have seen that holding the vowel constant and changing only the transitions, from those appropriate to [s] to those for [š], can shift the s/š boundary by as much as 200 to 500 Hz. In addition, a following high vowel produces a shift of around 200 to 500 Hz as well. When the noise is synthetic and the vocalic portion natural, as in the first experiment, the natural cues (here including transitions) appear to take on an accentuated importance. Thus the effects of both the transition and the vowel are larger in the first than in the second all-synthetic experiment.

The contribution of the transitions to the perception of the fricatives may be compared to their corresponding role in the perception of stops. Dorman, Studdert-Kennedy, and Raphael (1977) found that, in some contexts, the transitions had more effect on perception of place of articulation of the stop than the plosive burst; in other contexts, the burst was the deciding cue. Since the burst consists of noise produced at the point of articulation of the consonant, it has a natural analog in the noise of the fricative consonant, which is also produced at the point of articulatory constriction. When the fricative noise is a strong cue, as it is when it has the archetypal frequency values of [s] or [š], the noise is the deciding cue. If the noise is a weaker cue, as with the ambiguous noises, the transitions take on more weight as cues. The fricative noise is a robust signal, and the archetypal fricative noises are easily identifiable in isolation. It is interesting, therefore, that such a long and steady noise should show context effects similar to those of the much more transient stop burst.

The previously mentioned experiment by Kunisaki and Fujisaki (1977), which examined the effect of vowel context on the s/š boundary in Japanese, led to the hypothesis that the effect was due to the presence or absence of rounding. The present results do not directly address the question of whether rounding is the only relevant parameter. Still, the explanation that Kunisaki and Fujisaki offer for the effect of rounding does seem to apply, at least to some of the results, including the second experiment, if the synthetic vowels truly model the vowels they were intended to reproduce. Since anticipatory rounding in articulation (see Carney & Moll, 1971, Bell-Berti & Harris, 1979) would lower the frequency of sounds being produced, the appearance of a lower s/š boundary for the rounded vowels compared with the unrounded vowels ([u]

with [ï] and [ü] with [i]) would indicate that this coarticulation is being compensated for in perception. This holds for the comparison of [i] with [u] as well. If we were to take the same approach to the possible effects of the front/back distinction, then we would expect back vowels to have a lower boundary than front vowels. The resonating cavity should be somewhat longer for back vowels (based on the relatively retracted position of the tongue in producing [s] before [u] compared with [i] as seen in Carney and Moll, 1971), and the fricative noise would thus be somewhat lower as well. Thus the combined effect of the two dimensions would predict that [i] would give a high s/š boundary, [u] a low one, and [ï] and [ü], something in between. This is just what Experiment 2 gives us; however, Experiment 1 does not give clean enough results to make the comparison. More research is necessary to obtain a definitive explanation of the combined effects of the rounded/unrounded and front/back dimensions.

The transitions used in the second experiment did not exert as much influence on the fricative boundary as the natural transitions of Experiment 1. While the parallel decrease in the size of the vowel quality effect suggests that the vocalic segment as a whole provides a weaker set of cues when it is synthetic, the decrease could also be due to the presence in the natural speech of secondary cues that are not modelled in synthesis. In addition, the trajectories of the synthetic transitions were derived from a simple locus theory, which assumed one constant locus for each of the two fricatives. Cinefluorographic studies (Carney & Moll, 1971) show that this might not in fact be the case; the loci might shift before the different vowels. So the second test might not have reproduced as much of the coarticulation as the first test, which combined synthetic noises with natural transitions (which presumably would show this vowel-specific shift if it is really present).

Finally, there was no significant difference in the performance of the phonetically naive subjects compared to those with some linguistic sophistication and those with a good deal of sophistication. Whatever the effect of the following vowel turns out to be when more data is available, it seems that it is an effect which is shown even by those who do not use certain of the vowels distinctively. Thus, it seems that subjects are reacting to the acoustic or narrowly defined phonetic characteristics of the vowels rather than to their perceived (phonemic) category.

In sum, both the nature of the following vowel and its initial formant transitions contribute to the perception of fricative noises. The effect of the transitions is as large as, or larger than, the effect of the vowel. It seems that the vowel effect is insensitive to linguistic experience.

REFERENCES

Bell-Berti, F., & Harris, K. S. Anticipatory coarticulation: Some implications from a study of lip rounding. Journal of the Acoustical Society of America, 1979, 65, 1268-1270.

Carney, P. J., & Moll, K. L. A cinefluorographic investigation of fricative consonant-vowel coarticulation. Phonetica, 1971, 23, 193-202.

Delattre, P. C., Liberman, A. M., & Cooper, F. S. Formant transitions and loci as acoustic correlates of place of articulation in American fricatives. *Studia Linguistica*, 1964, *18*, 104-121.

Dorman, M. F., Studdert-Kennedy, M., & Raphael, L. J. Stop-consonant recognition: Release bursts and formant transitions as functionally equivalent, context-dependent cues. *Perception & Psychophysics*, 1977, *22*, 109-122.

Harris, K. S. Cues for the discrimination of American English fricatives in spoken syllables. *Language and Speech*, 1958, *1*, 1-7.

Kunisaki, O., & Fujisaki, H. On the influence of context upon perception of voiceless fricative consonants. *Annual Bulletin* (Research Institute of Logopedics and Phoniatrics, University of Tokyo), 1977, *11*, 85-91.

LaRiviere, C., Winitz, H., & Herriman, E. The distribution of perceptual cues in English prevocalic fricatives. *Journal of Speech and Hearing Research*, 1975, *18*, 613-622.

McCasland, G. P. Formant transition place cues of intervocalic fricatives. *Journal of the Acoustical Society of America* (Supplement No. 1), 1978, *63*, S21. (Abstract)

INFLUENCE OF VOCALIC CONTEXT ON PERCEPTION OF THE [ʃ]-[s] DISTINCTION: I. TEMPORAL FACTORS

Virginia A. Mann and Bruno H. Repp

Abstract. When synthetic fricative noises from an [ʃ]-[s] continuum are followed by [a] or [u] (with appropriate formant transitions), listeners perceive more instances of [s] in the context of [u] than in the context of [a]. Presumably, this reflects a perceptual adjustment for the coarticulatory effect of rounded vowels on preceding fricatives (through anticipatory lip rounding). We replicated the basic perceptual effect and collected acoustic data from one speaker to corroborate the presence of an analogous coarticulatory effect in production. We also found that varying the duration of the fricative noise leaves the perceptual effect unchanged, whereas insertion of a silent interval following the noise reduces it substantially. Subsequently, we tried to determine whether it is mere temporal separation or the perception of an intervening stop consonant that is responsible for this reduction. The results suggest temporal separation as the important factor, which agrees with recent, analogous observations on anticipatory lip rounding.

INTRODUCTION

Acoustic analyses of speech have revealed that the noise spectrum of fricative consonants varies with the nature of the following vowel (Bondarko, 1969; Fujisaki & Kunisaki, 1978). This acoustic context dependency seems to be primarily, although not exclusively, a consequence of anticipatory lip rounding for vowels such as [u] and [o], which results in a lowering of the fricative noise spectrum. Zue (Note 1) has demonstrated analogous variations in the spectrum of stop consonant bursts with the following vowel.

This coarticulatory effect has a parallel in speech perception: Listeners' identifications of fricative consonants are influenced by vocalic context. Evidence for such a dependency has been scattered through the literature for some time (Delattre, Liberman, & Cooper, 1962; Hughes & Halle,

---

Acknowledgment: This research was supported by NICHD Grant HD01994 and BRS Grant RR05596 to the Haskins Laboratories, and by NICHD Postdoctoral Training Fellowship HD05677 to Virginia Mann. Some of the results were reported at the 96th Meeting of the Acoustical Society of America in Honolulu, December 1978, and at the 10th International Congress of Phonetic Sciences in Copenhagen, August 1979. We thank Patti Price and Sarah Peck for their help in conducting and analyzing the experiments, and Alvin Liberman and David Isenberg for helpful comments on an earlier draft of this paper.

[HASKINS LABORATORIES: Status Report on Speech Research SR-59/60 (1979)]

1956; Hasegawa & Daniloff, Note 2), but the clearest demonstration was provided in a recent study by Kunisaki and Fujisaki (1977). Using a continuum of synthetic fricative noises varying from [ʃ] (low pole frequencies) to [s] (high pole frequencies), these researchers found that the category boundary shifted in favor of [s] responses when the noises were followed by synthetic [u] or [o], relative to the boundaries obtained in the context of [a] or [e]. In other words, the phoneme boundary shifted toward lower noise frequencies in the context of rounded vowels, in conformity with the analogous effect of anticipatory lip rounding on fricative noise spectra. Thus, the (Japanese) listeners seemed to take account in perception of contextual changes characteristic of fricative production.

It is intriguing to assume that listeners have an intrinsic knowledge of articulatory dynamics, and that their phonetic perception is guided by this knowledge. However, speech perception necessarily makes use of the mechanisms of auditory perception, and there are a variety of psychoacoustic factors that may interact with--or, indeed, prevent--the link between perception and production that presumably underlies speech perception at the highest level. With this in mind, we have investigated some of the temporal and spectral stimulus parameters that influence (and sometimes limit) the effects of vocalic context on the perception of fricatives. In the present paper, we will be concerned with temporal parameters; in a subsequent paper (Repp & Mann, Note 3), we will report our investigations of spectral stimulus properties. By determining the roles played by these parameters, we hoped to constrain the possible psychoacoustic explanations of the perceptual context effect. Furthermore, future investigations of analogous parameters in speech production should enable us to draw a closer parallel between perception and production of fricatives.

EXPERIMENT I

The purpose of our first experiment was to replicate the basic finding of Kunisaki and Fujisaki (1977) that the phonetic perception of a fricative noise depends on the nature of the following vowel, and then to determine how the magnitude of that perceptual context effect changes as a function of two variables: the duration of the fricative noise and the presence or absence of a silent interval between the noise and the vocalic portion. While changes in noise duration, within the range employed by us, have no gross effect on phonetic perception, insertion of a silent interval induces perception of a stop consonant (cf. Bastian, Eimas, & Liberman, 1961; Bailey & Summerfield, in press) and thus changes the phonetic structure of the stimulus. Nevertheless, there was no a priori reason to assume that either of these two temporal manipulations would be more effective than the other in reducing the contextual effect of the vowel on the fricative. If listeners assign a phonetic category to the fricative as soon as some of the noise has been processed, then the temporal distance between noise onset and vocalic portion should be the important variable, and it should not matter whether this distance is varied by extending the duration of the noise or by inserting a silent interval after it.

Method

Subjects. The subjects included nine paid student volunteers recruited from Yale University, one research assistant, and the two investigators. With the exception of the second author, no subject had extensive experience in listening to synthetic speech, although some had participated in earlier experiments of a similar nature. All but two of the subjects were native speakers of American English; the remaining two were native speakers of German and Chinese, respectively, but fluent in English. Neither experience nor language seemed to affect the pattern of results, so the data of all 12 subjects were combined.

Stimuli. A synthetic fricative noise continuum was created on the OVEIIIc serial resonance synthesizer at Haskins Laboratories, following in part the specifications given by Kunisaki and Fujisaki (1977). Each noise was characterized by two steady-state poles (formants) produced by the fricative circuit of the synthesizer. No zero (antiformant) was specified. There were nine different stimuli. The center frequencies of both poles increased from stimulus 1 ([ʃ]-like) to stimulus 9 ([s]-like) in roughly equal steps; the step size was larger for the second (higher) pole than for the first. These frequencies are listed in Table 1. Each noise reached full amplitude after 40 msec and decreased in amplitude over the last 30 msec. Due to the characteristics of the synthesizer, which are intended to mimic natural speech, the noise amplitude increased from stimulus 1 to stimulus 9 by approximately 12 dB. This characteristic of the stimuli was left intact. Stimulus duration was 100 or 250 msec, depending on the condition.

Table 1

Pole frequencies of fricative noises (Hz).

Stimulus	Pole 1	Pole 2
1	1957	3803
2	2197	3915
3	2466	4148
4	2690	4269
5	2933	4394
6	3199	4655
7	3389	4792
8	3591	4932
9	3917	5077

In addition to the fricative noise continuum, we synthesized two vocalic stimuli with initial formant transitions roughly appropriate for an alveolar voiceless unaspirated stop consonant: [ta] and [tu]. (The formant transitions, which approximated those normally following [s] and [ʃ], were required to make the fricative noise and vocalic portions perceptually coherent; see

Repp and Mann, Note 3.) Each of these stimuli was 200 msec in duration, with a 70-msec amplitude ramp at onset, and a fundamental frequency contour that fell linearly from 110 to 80 Hz. The steady-state frequencies of the first three formants were 771, 1233, and 2520 Hz for [a], and 250, 800, and 2295 Hz for [u]. [ta] had 50-msec stepwise-linear transitions in the first and second formants with starting frequencies of 500 and 1796 Hz, respectively. [tu] had a 70-msec stepwise-linear transition in the second formant only, with a starting frequency of 1499 Hz. The amplitudes of [ta] and [tu] were adjusted to be approximately equal.[1] They were 12-24 dB higher than those of the fricative noises which, as mentioned above, varied over a 12-dB range.

The experiment had five conditions, distinguished by the composition of the stimuli:

(1) Isolated 250-msec noises.

(2) Short (100-msec) noises, immediately followed by either [ta] or [tu].

(3) Long (250-msec) noises, immediately followed by either [ta] or [tu].

(4) Short (100-msec) noises, followed by a 150-msec silent gap and either [ta] or [tu].

(5) Long (250-msec) noises, followed by a 150-msec silent gap and either [ta] or [tu].

As can be seen, conditions 2-5 represented the factorial combination of two variables: noise duration (100 or 250 msec) and gap duration (0 or 150 msec). In conditions 2 and 3, listeners did not perceive any stop consonants because there was no silence indicating closure. Thus, listeners heard reasonable instances of [ʃa], [sa], [ʃu], and [su]. In conditions 4 and 5, there was a gap of more than sufficient duration to enable listeners to hear a stop consonant; thus, [ʃta], [sta], [ʃtu], and [stu] were perceived. Although [ʃt] clusters do not occur in initial position in English, they appeared to pose no perceptual difficulty for our listeners.

All stimulus sequences were recorded directly from the synthesizer onto magnetic tape. Condition 1 contained three random sequences of 42 stimuli each, with interstimulus intervals (ISIs) of 3 sec, and 6 sec between sequences. The other four conditions each contained five such sequences. In all conditions, the nine stimuli from the fricative noise continuum occurred with unequal frequencies according to a 1-2-3-3-3-3-3-2-1 schedule, leading to a basic set of 21 stimuli. This set was replicated once within each sequence in condition 1, whereas in the other conditions the two different vocalic portions, [ta] and [tu], led to 42 stimuli in each sequence. All in all, each listener gave 15 responses (18 in condition 1) to each of the more ambiguous fricative noises (stimuli 3-7 on the continuum).

Procedure. The five conditions were presented in the same fixed order (1-5) to all subjects, with brief pauses in between. Subjects were seated in a quiet room and listened over Telephonics TDH-39 earphones at a comfortable intensity. The tapes were played back on an Ampex AG-500 tape recorder. The task was the same in all conditions--to identify in writing the fricative

61

consonant in each stimulus as either "sh" or "s."

Results

The results of this experiment are shown in Figure 1. Consider first the dotted function connecting the triangles in Figure 1b. (The function is duplicated in Figure 1d.) It represents the percentage of "sh" responses to the nine isolated noises (condition 1). It can be seen that all listeners reliably identified the end points of the noise continuum as "sh" and "s", respectively. Stimuli 3-7 showed varying amounts of ambiguity, but there was a reasonably orderly progression from "sh" to "s" responses. With the exception of one subject who gave rather inconsistent responses, all individual category boundaries fell in the vicinity of stimulus 5 (mean = 5.22; standard deviation = 0.41), indicating relatively little variation in response criteria between listeners.

Figures 1a and 1b show the effect of immediately following the fricative noises with a vocalic portion. It can be seen that the predicted effect of vocalic context was obtained: Listeners reported more "sh" sounds when [(t)a] followed than when [(t)u] followed. (The parentheses indicate that [t] was not actually perceived.) This effect, which replicates Kunisaki and Fujisaki (1977), was obviously very large and included even stimuli at the [ʃ]-end of the continuum. Comparison with the baseline results for isolated noises (Figure 1b) shows that the context effect was primarly due to [(t)u] which pulled the level of "sh" responses down. This is exactly what was to be expected if the perceptual effect of vowel context parallels the coarticulatory effect of anticipatory lip rounding. Since [(t)a] does not involve lip rounding, this context would not be expected to shift responses from the baseline level.

Comparison of Figures 1a and 1b indicates that extending the duration of the fricative noise from 100 to 250 msec left the context effect virtually unchanged. On the other hand, a glance at Figures 1c and 1d shows that the introduction of a 150-msec gap between the noise and the vocalic portion practically eliminated the effect. Note that conditions 2 and 3 (Figures 1b and 1c) represent the same interval (250 msec) between noise onset and onset of periodicity; however, in one case the first 100 msec of the noise were followed by more noise whereas silence followed in the other case. Clearly, the silent interval in condition 3 had a different effect on perception than the noise-filled interval in condition 2. There was also an indication of a slight overall decrease in "sh" responses (relative to the baseline) in condition 4 (Figure 1c). This may have been due to the short duration of the noises.

The statistical analysis confirmed these observations. A three-way analysis of variance was conducted on the response percentages summed over all noise stimuli, with the factors Context, Noise Duration, and Gap. Context had a highly significant effect, F(1,11) = 55.7, p < .001, and this effect interacted with Gap, F(1,11) = 62.5, p < .001, but not with Noise Duration, F(1,11) = 1.6. In addition, there was a main effect of Noise Duration, F(1,11) = 12.0, p < .01, and an interaction of this factor with Gap, F(1,11) = 7.0, p < .025, both effects being due to the decrease in "sh" responses in condition 4 (short noise plus gap).

Separate analyses of conditions 2 and 3 and conditions 4 and 5 confirmed that fricative noise duration had no significant effect on the category boundary shift, regardless of whether a gap was present or not. However, reducing the duration of the noise significantly decreased the number of "sh" responses when the 150-msec gap was present, $F(1,11) = 23.2$, $p < .001$, but not in the absence of a gap, $F(1,11) = 0.5$. Interestingly, the vowel context effect at the 150-msec gap, though small, was still highly significant, $F(1,11) = 17.6$, $p < .01$. Thus, although the introduction of the gap substantially reduced the context effect, it did not completely eliminate it.

Discussion

Our results partially replicate the findings of Kunisaki and Fujisaki (1977) on the effects of vocalic context on the perception of the distinction between [ʃ] and [s]. Although their data were presented in a somewhat different format, some comparison parameters can be derived from their figures. There was a striking difference in absolute boundary locations between their listeners and ours. Expressing boundary locations in terms of first-pole frequency, we find average boundary locations for our listeners at approximately 2570 and 3060 Hz in [-(t)u] and [-(t)a] context, respectively (cf. Fig. 1a and Table 1), whereas Japanese subjects showed corresponding boundaries at approximately 3200 and 3900 Hz. This large difference suggests language-specific differences in the [ʃ]-[s] distinction.[2] Moreover, it is evident that the Japanese listeners showed a larger context effect (about 700 Hz) than our subjects (about 500 Hz). However, there were enough changes in detailed stimulus structure and method between their study and ours to account for this difference.

Next, we may ask whether the magnitude of the fricative boundary shift as a function of vocalic context corresponds to the magnitude of the analogous spectral shifts in production. Kunisaki and Fujisaki (1977) report acoustic measurements for Japanese. There, the average shift in first-pole frequency between [-a] and [-u] contexts was about 100 Hz for [ʃ] and 300 Hz for [s], although there was considerable variability. Surprisingly, these differences are considerably *smaller* than the perceptual boundary shifts obtained in the Japanese study (about 700 Hz).

We have been unable to find in the literature systematic spectral measurements of American English fricative noises in the vocalic contexts that we employed. To get some preliminary impression of the magnitude of the coarticulatory effect, we recorded a male native speaker of American English saying [ʃa], [ʃu], [sa], and [su] twelve times in random order. Subsequently, we digitized these utterances at 10 kHz (using the Haskins Laboratories Pulse Code Modulation system) and examined successive spectral cross-sections (12.8 msec time window) of the fricative noises. In each spectrum, we measured the frequency of the lowest peak (which may or may not have represented the first pole) and subsequently averaged these measurements over all cross-sections of a given token. Finally, we averaged over the 12 tokens of each utterance. These means and the associated standard deviations are shown in Table 2. It is evident that, for this single speaker at least, the first spectral peak was lowered by about 250 Hz in [-u] context, relative to [-a] context. This shift is only about half the size of the perceptual boundary shift found for American listeners, which confirms our similar observation on Japanese lis-

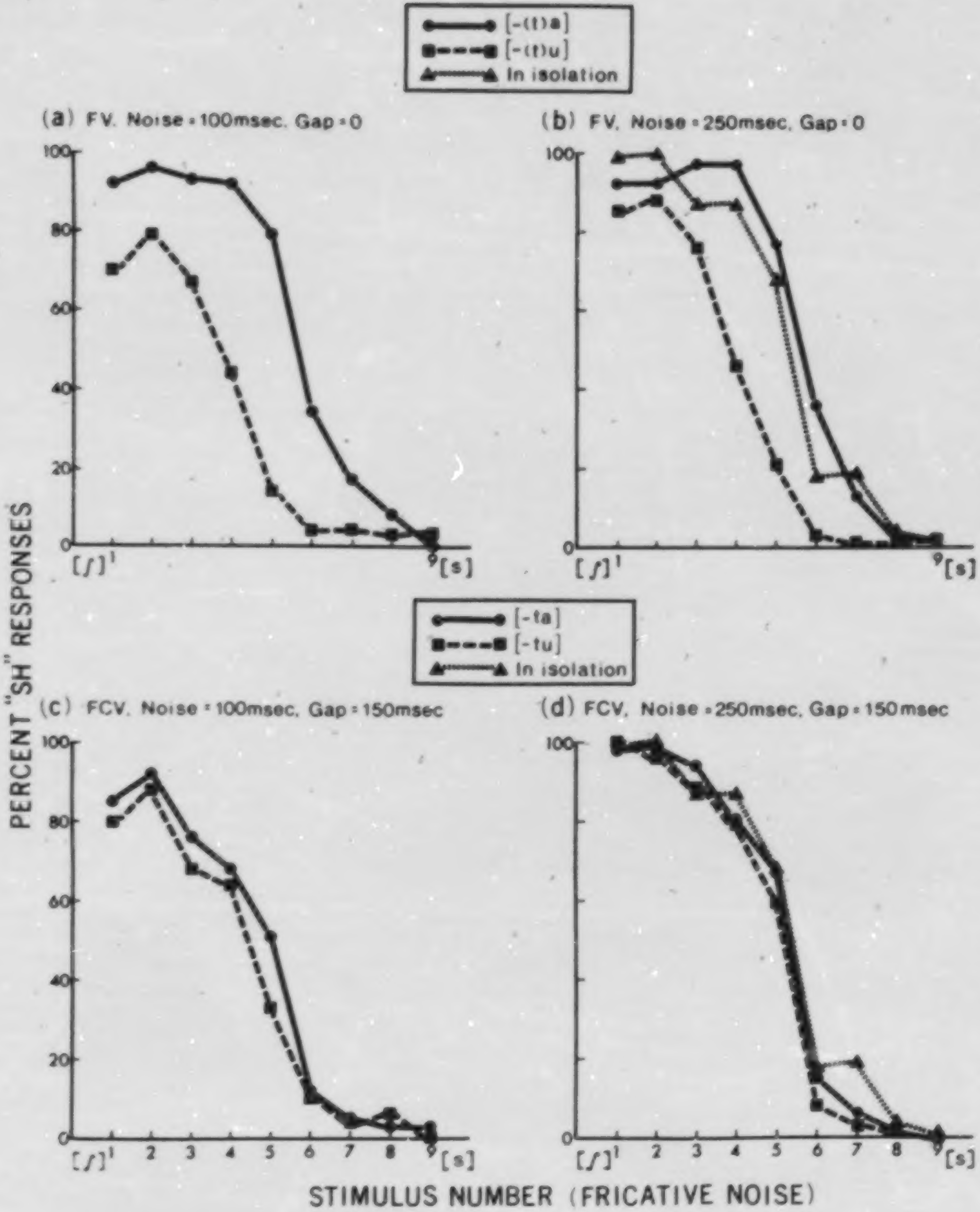


Figure 1. Effect of vocalic context on the [ʃ]-[s] contrast in four conditions.

teners. Thus, these comparisons suggest that listeners' intrinsic knowledge of coarticulatory effects in production is not the only factor affecting perception, or that there is perceptual overcompensation.

Table 2

Acoustic context effects of [a] and [u] on preceding [ʃ] and [s].

Utterance	Frequency of first spectral peak in fricative noise (Hz)	
	Mean (12 tokens)	Standard deviation
[ʃa]	2405	94
[ʃu]	2115	133
[sa]	3773	149
[su]	3563	116

The present study extended the Kunisaki-Fujisaki study by examining the effects of two temporal variables on the magnitude of the perceptual boundary shift. We found little effect of a change in the duration of the fricative noise from 100 to 250 msec--a range of durations which exceeds that of normal fricative noises in running speech (cf. Klatt, 1974; Umeda, 1977). This suggests that the critical perceptual information is located at the end of the fricative noise, where it adjoins the vocalic portion, rather than at its onset. The finding that the introduction of a silent gap between the noise and the periodic portion nearly eliminated the effect further demonstrates that the perceptual interaction of the two stimulus portions may depend on their temporal contiguity. This is reasonable, since anticipatory lip rounding in production, if it is not fully established prior to the onset of the utterance, would be expected to affect the later portion of the fricative noise more than the earlier portions (Bondarko, 1969). Moreover, Bell-Berti and Harris (1979) have recently claimed that the onset of lip rounding precedes a rounded vowel by a certain fixed time interval. If this is true, it might also be the case that the fricative noise must fall within a certain distance from the vowel in order for its perception to exhibit a contextual influence.

There seems little point in further investigating the variable of noise duration. Given that a 250-msec noise is already beyond the range of durations normally encountered in running speech, extending noise duration further, even though it might eventually lead to a decline of the vowel context effect, would provide data of little relevance to the perception of speech. However, the question of how much separation between noise and periodic portions is needed to prevent contextual effects is of greater theoretical interest. This is so because an additional factor may play a role: the perception of an intervening stop consonant when silence is introduced. Is it temporal separation *per se* that reduces the contextual effect, or is it the perception of an intervening phonetic segment? Experiment II was designed to answer that question.

EXPERIMENT II

Before conducting an experiment that systematically varied gap size, we collected data for stimuli with a gap duration of 75 msec--halfway between the gap sizes used in Experiment I, and more than enough for a stop consonant to be heard. The duration of the fricative noise in these stimuli was 150 msec. The stimulus sequence was similar to those of conditions 2-5 in Experiment I, and the same 12 subjects listened to it in a separate session. The results showed a highly significant context effect, $F(1,11) = 93.5$, $p < .0001$, which was nevertheless rather small, similar to that obtained with a 150-msec gap duration (Figure 1d).[3] Indeed, the difference between the 75-msec and 150-msec gap conditions fell short of significance in a separate test, $F(1,11) = 4.1$, $p > .05$, and both effects were much smaller than that obtained with no gap at all. Since these data suggested a major decrease in the vowel context effect at gap durations shorter than 75 msec, we decided to focus our attention on these short intervals.

Method

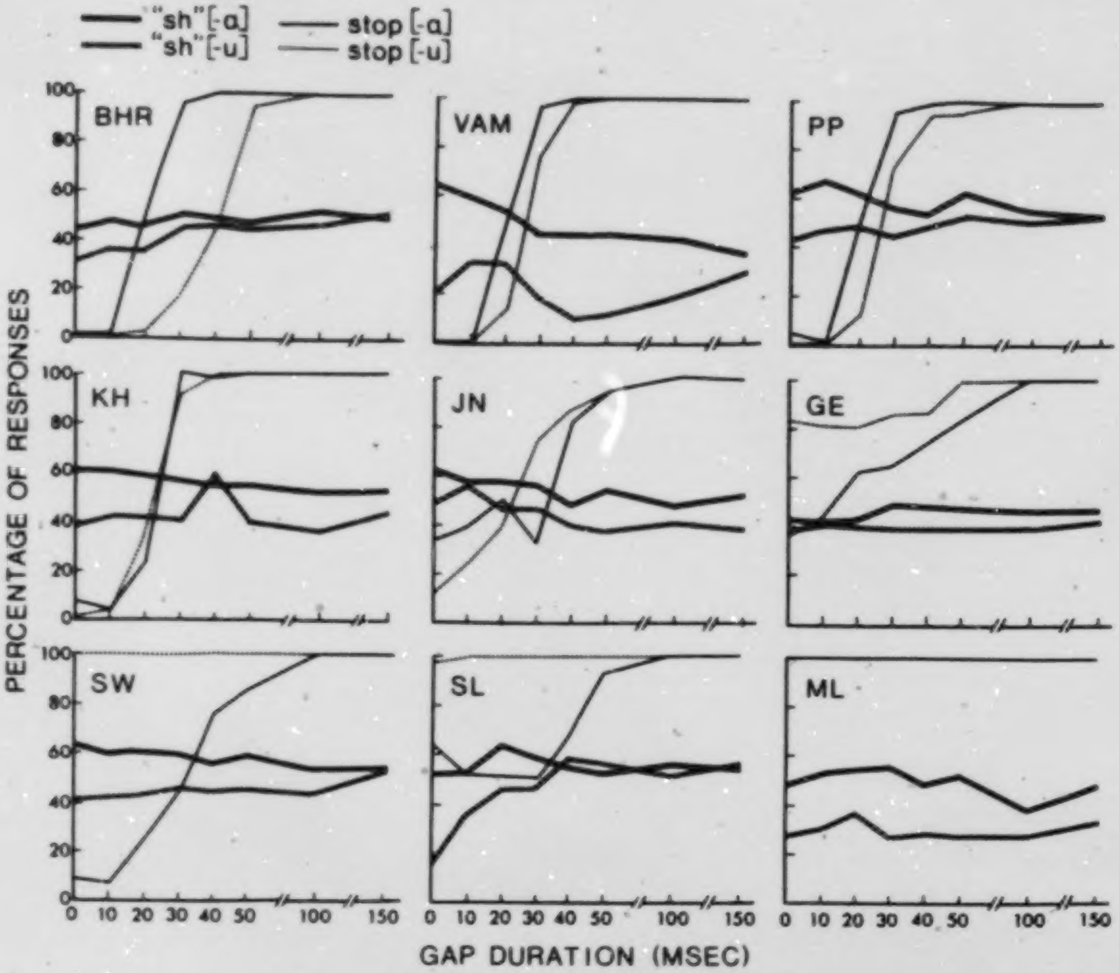
Subjects. Nine subjects participated in this experiment, including seven paid volunteers, a research assistant, and the two investigators. Half of one subject's data were rejected since he gave so few "s" responses in the first session.

Stimuli. The stimuli were similar to those used in Experiment I. One change concerned the amplitudes of the fricative noises. In Experiment I, noise amplitude increased strongly from [ʃ] to [s]. Although that inequality had been built into the synthesizer by its manufacturer, presumably to model natural speech, it seemed somewhat extreme to us and, moreover, was not in accord with observations of our own gathered in the meantime (cf. Repp & Mann, Note 3). We therefore modified the amplitude settings in the synthesizer parameters so as to make all fricative noises approximately equal in amplitude. This was achieved by specifying lower amplitudes for the more [s]-like noises, resulting in a relatively constant amplitude difference between fricative noise and vocalic portion of about 24 dB. The fricative noises were 150 msec long and had 50-msec initial and final amplitude ramps. There were eight gap durations: 0, 10, 20, 30, 40, 50, 100, and 150 msec.

The tape recorded for the experiment contained three random sequences of 144 stimuli, separated by 3-sec ISIs. Each sequence contained the 18 combinations of the nine fricative noises with [ta] and [tu], at each of the eight gap sizes. In contrast to Experiment I, all nine fricative noises occurred with equal frequency. Gap durations were totally randomized in the test sequences.

Procedure. Each subject listened to the experimental tape four times, in two separate sessions. The task was to identify both the fricative and the stop consonant (if present) in each stimulus. The response choices were "s," "sh," "st," "sht," "sk" and "shk."[4]

Figure 2. Effect of vocalic context on the [ʃ]-[s] distinction as a function of silent gap duration and stop consonant perception. Individual subject data from Experiment II.



Results

Assuming that the basic vowel context effect would be replicated when no silence intervened between the noise and the vocalic portion, we expected the context effect to exhibit a sharp decline as gaps of increasing duration were inserted. The form of this decline was of special interest: Would it be continuous with increases in gap duration, or would it show a discontinuity at the point where stop consonants began to be heard?

The results are shown in Figure 2. The data are displayed separately for the nine subjects, in order to show the considerable individual differences. Each of the nine panels contains four response functions: The two steeply rising ones (thin lines) represent the increase in the percentage of stop responses in [-a] and [-u] context as gap duration was increased; the two more nearly horizontal functions (heavy lines) represent the percentage of "sh" responses (averaged over the whole fricative-noise continuum) in [-a] and [-u] context. The difference between the latter two functions is a measure of the magnitude of the vowel context effect, with a 10-percent difference representing a category boundary shift of roughly 200 Hz on the first-pole-frequency dimension of the synthetic noises.

First of all, it is evident that the basic context effect was indeed replicated: All subjects gave more "sh" responses in the [-(t)a] context than in the [-(t)u] context, $F(1,8) = 33.22$, $p < .001$. There was, however, considerable variability in both the magnitude of the effect, and in its relation to gap size. One subject (SL) showed a complete disappearance of the context effect at 40 msec of silence; two other subjects (BHR and PP) showed a progressive reduction up to that interval. The remaining subjects showed little change in the magnitude of context effect for gap sizes up to 50 msec. Analysis of variance of the 0-50 msec gaps revealed only a marginally significant and slightly irregular overall decline in the context effect with gap duration, $F(5,40) = 3.31$, $p < .05$. Evidence for a decline of the context effect at longer gap sizes was more convincing; it was significant in an analysis of variance including the 0, 50, 100, and 150 msec intervals, $F(3,24) = 8.54$, $p < .001$. Nevertheless, at least three subjects still exhibited sizeable context effects at the longest gap duration, 150 msec.

We turn now to stop consonant perception as a function of gap duration, in order to address the question of whether the perception of an intervening stop limits the occurrence of a context effect between vowel and fricative noise. The stop/no-stop boundaries for four of the nine subjects (BHR, VAM, PP, KH) were quite regular: No stop consonants were heard at the shortest gap durations (0, 10 msec), and 30-40 msec of silence were sufficient to hear stops in most cases. Three of these subjects heard stops earlier (i.e., at shorter gap durations) in [-a] context than in [-u] context, a finding which is in agreement with results obtained by Bailey and Summerfield (in press).5 The responses of the remaining five subjects were more irregular. One of them (ML) heard stops at all gap durations, including stimuli without any true silence at all. Three subjects (SW, SL, GE) heard stops in all (or nearly all) [-u] stimuli, regardless of gap size, although they tended to hear no stops in [-a] context on at least some trials when gap duration was short. (Note that this context effect runs counter to that for the four subjects with a more regular response pattern.) The remaining subject (JN) showed no such context effect but a moderate tendency to hear stops even at short gap durations.[6] Despite this striking variability in the onset of stop percepts, the data provide clear evidence against the hypothesis that the perception of a stop consonant blocks the effect of a following vowel on fricative perception. Inspection of Figure 2 reveals that the onset of stop consonant perception generally is not accompanied by a marked reduction in the magnitude of the context effect. The only possible exception is subject SL for whom the context effect disappeared as soon as all stimuli were perceived as containing stops. However, this subject (and others as well) showed a large context effect at short gap durations despite a strong tendency to hear stops, which in itself argues against an inhibitory role of stop percepts.[7]

Discussion

The results of Experiment II justify the conclusion that the perception of an intervening stop consonant has relatively little influence on the effect of vocalic context on fricative labeling. For a few subjects, this context effect may have been slightly reduced by the perception of an intervening segment; however, the majority of subjects remained unaffected and showed only a slow decline of the context effect with increasing temporal separation of fricative noise and vocalic portion. In some cases, the context effect seemed to extend across more than 150 msec of silence.

To the extent that temporal separation was more important than the number of phonetic segments perceived, the present results are in agreement with the speech production data of Bell-Berti and Harris (1979). However, the large individual differences and the temporal extent of the context effect for some listeners suggest that it may be difficult to compare temporal parameters between speech perception and production. In perception, and perhaps in production as well, individual strategies may modify whatever basic, underlying phenomenon there may be. In the present case, for example, individually varying tendencies to perceive the fricative noise either as forming a unit with the vocalic portion or as "streaming off" as a separate auditory event may have played a role. Perhaps the context effect could be extended over arbitrarily long temporal separations if listeners made an effort to integrate the fricative and CV portions into a single perceptual unit. The individual differences observed in the present study may in part have derived from differences in strategies of perceptual integration.

The results of Experiment II speak to a question that we address more directly in experiments reported in a separate paper (Repp & Mann, Note 3): Is the context effect of the vocalic portion on the fricative indeed due to vowel quality itself--as we have assumed all along--or is it is perhaps due, in part or entirely, to the initial formant transitions of the vocalic portions? Although, following the seminal study of Harris (1958), vocalic formant transitions were believed to be unimportant for the [ʃ]-[s] contrast, recent experiments by Whalen (1979) show that the transitions are a strong cue when the fricative noise is ambiguous (cf. also Delattre et al., 1962, for similar results on voiced fricatives). The formant transitions of [ta] and [tu] in the present experiments were chosen on the basis of the experimenters' intuitions, and it could have been the case that one set of transitions favored "s" (or "sh") percepts more than did the other. However, two observations suggest that the context effects observed in Experiments I and II were largely due to vowel quality. First, it seems reasonable to argue that, as soon as the formant transitions are interpreted as a cue to place of articulation of an intervening stop consonant (rather than of the initial fricative), they should lose whatever effect they might have had on fricative perception when no stop was heard. If this hypothesis is correct, then any context effects that are observed despite an intervening stop percept--and Experiment II provided ample evidence for such effects--must be due to vowel quality alone.

Second, and more importantly, if the context effect--especially at short gap durations--had been due to the formant transitions in [ta] and [tu] acting as cues to fricative place of articulation, then the transitions of [tu] should have been more appropriate for a forward (dental) place of articulation (thus favoring "s" percepts) than those of [ta] (which favored "sh" percepts, i.e., an alveolar place of articulation). Although both stimuli in isolation sounded to us as beginning with a "t," many subjects gave a substantial proportion of "k" responses when the same stimuli were preceded by a fricative noise plus a sufficient amount of silence to permit perception of a stop. According to the argument just made, "k" responses should have been less frequent with [tu] than with [ta] if the transitions of [tu] favored a more forward place of articulation ("t" responses, in this instance). In fact, the opposite was observed. Of the nine subjects, seven gave "k" responses only or predominantly to our [tu]; one subject showed little difference between [ta] and [tu]; and only one subject (SL) showed the opposite pattern, giving "k" responses to [ta] only.[8]

Thus, it seems that, for the large majority of the subjects, the context effect must have been due to vowel quality, even at short gap durations. In fact, if the transitions indeed contributed to fricative perception at short gap durations (when no stop was heard), the transition effect may have partially cancelled the vowel quality effect in these subjects, and this may have been the reason why the reduction in the overall context effect at the stop/no-stop boundary was not more pronounced. In order to investigate this possibility, it will be necessary to dissociate the transition and vowel quality effects experimentally, and then to examine the influence of systematic variations in gap size on the two separate context effects. In a separate paper (Repp & Mann, Note 3), we report experiments that achieve such a dissociation (see also Whalen, 1979) and demonstrate independent effects of both transitions and vowel quality on fricative perception. However, we do not yet know exactly how these two separate effects change with variations in gap size. Until we have this information, our conclusion that the vocalic context effect is unaffected by an intervening phonetic segment must remain tentative. Certainly, however, perception of an intervening stop consonant does not prevent effects of vocalic context on fricative perception. Our conclusion stands that temporal separation is the primary factor affecting the size of the context effect investigated here.

REFERENCE NOTES

1. Zue, V. W. Acoustic characteristics of stop consonants: A controlled study. *Lincoln Laboratory Technical Report No. 523*. Lexington, MA, 1976.
2. Hasegawa, A., & Daniloff, R. G. *Effects of vowel context upon labelling*

the /s/-/š/ continuum. Paper presented at the 91st Meeting of the Acoustical Society of America, Washington, D.C., April 1976.

3. Repp, B. H., & Mann, V. A. *Influence of vocalic context on perception of the [ʃ]-[s] distinction: II. Spectral factors*. Manuscript, 1979.

REFERENCES

Bailey, P. J., & Summerfield, Q. Some observations on the perception of [s]+stop clusters. *Journal of Experimental Psychology: Human Perception and Performance*, in press.

Bastian, J., Eimas, P. D., & Liberman, A. M. Identification and discrimination of a phonemic contrast induced by silent interval. *Journal of the Acoustical Society of America*, 1961, *33*, 842. (Abstract)

Bell-Berti, F., & Harris, K. S. Anticipatory coarticulation: Some implications from a study of lip rounding. *Journal of the Acoustical Society of America*, 1979, *65*, 1268-1270.

Bondarko, L. V. The syllable structure of speech and distinctive features of phonemes. *Phonetica*, 1969, *20*, 1-40.

Delattre, P. C., Liberman, A. M., & Cooper, F. S. Formant transitions and loci as acoustic correlates of place of articulation in American fricatives. *Studia Linguistica*, 1962, *16*, 104-121.

Fujisaki, H., & Kunisaki, O. Analysis, recognition, and perception of voiceless fricative consonants in Japanese. *IEEE Transactions (ASSP)*, 1978, *26*, 21-27.

Harris, K. S. Cues for the discrimination of American English fricatives in spoken syllables. *Language and Speech*, 1958, *1*, 1-7.

Hattori, S. Phoneme, phone, and compound phone. In S. Hattori, *Gengo no hoho*. Tokyo: Iwanami, 1960, 751-763.

Heinz, J. M., & Stevens, K. N. On the properties of voiceless fricative consonants. *Journal of the Acoustical Society of America*, 1961, *33*, 589-596.

Hughes, G. W., & Halle, M. Spectral properties of fricative consonants. *Journal of the Acoustical Society of America*, 1956, *28*, 303-310.

Klatt, D. H. The duration of [s] in English words. *Journal of Speech and Hearing Research*, 1974, *17*, 51-63.

Kunisaki, O., & Fujisaki, H. On the influence of context upon perception of voiceless fricative consonants. *Annual Bulletin of the Research Institute for Logopedics and Phoniatrics*, 1977, *11*, 85-91.

Mann, V. A., & Repp, B. H. Influence of preceding fricative on stop consonant perception. *Haskins Laboratories Status Report on Speech Research*, 1979, *SR-59/60*, this issue.

Umeda, N. Consonant duration in American English. *Journal of the Acoustical Society of America*, 1977, *61*, 846-858.

Whalen, D. H. Effects of vocalic formant transitions and vowel quality on the English /s/-/š/ boundary. *Haskins Laboratories Status Report on Speech Research*, 1979, *SR-59/60*, this issue.

FOOTNOTES

[1] This adjustment was made in the synthesis parameters. Given equal amplitude parameters, [ta] would have emerged from the synthesizer with

considerably higher amplitude than [tu]. Although this difference is intended to mimic natural speech, we found it undesirable to confound such a large amplitude difference (about 10 dB) with the effect of vowel context we were looking for. Thus, we chose the lesser evil of not preserving the natural amplitude relationships of [ta] and [tu]. Essentially, we believe that amplitude variations will have little influence on the context effect under study but, as yet, we have no data to support this prediction.

[2] It may be the case that [ʃ] and [s] in Japanese have different spectra (and correspondingly, different articulatory positions) than in English, causing native speakers of the two languages to place their perceptual boundaries differently, in accordance with their language experience. An articulatory difference is suggested especially for [ʃ], which by some Japanese linguists is considered a compound phone, [sj], equivalent to a palatalized [s] (e.g., Hattori, 1960). Kunisaki and Fujisaki (1977) and Fujisaki and Kunisaki (1978) report data indicating that the average frequencies for the first pole (formant) of Japanese [ʃ] and [s] lie at about 2800 Hz and 4000 Hz, respectively. The average Japanese perceptual boundary occurred around 3500 Hz on this dimension. On the other hand, Heinz and Stevens (1961) report average first-pole frequencies for American [ʃ] and [s] of approximately 2400 Hz and 4800 Hz, respectively, which suggests that the spectra of these fricatives are more distinct in American English than in Japanese, but provides no clue as to why the perceptual boundary is lower for American listeners (viz., at about 2800 Hz). There are various other factors that might have played a role: the stimulus range employed, the relative amplitudes of the fricative noises (cf. Heinz & Stevens, 1961), the nature of the formant transitions in the vocalic portions (cf. Harris, 1958), and the fact that no zeros (antiformants) were specified for the present fricative stimuli (cf. Fujisaki & Kunisaki, 1978). There were differences in all these respects between the Japanese study and ours, and it would lead too far to attempt to discuss each in detail. However, it should be noted that Hasegawa and Daniloff (Note 2) synthesized a [ʃ]-[s] continuum by a method rather similar to that of Kunisaki and Fujisaki (1977) and found, for American listeners, a perceptual boundary closer to ours, viz., at about 2700 Hz of first-pole frequency. Thus, cross-language differences in perception and production of these fricatives are indicated.

[3] The high level of significance of the 75-msec context effect was due to its remarkable consistency across subjects: All twelve listeners showed a small effect in the expected direction.

[4] When serving as subjects in Experiments I, we had noticed a tendency to hear velar stops on occasion, even though the periodic stimulus portions in isolation were heard as beginning with alveolar stops. Our informal observation that the tendency to hear velar stops was much stronger following [s] than following [ʃ] led to a series of separate studies of this phenomenon (Mann & Repp, 1979). The present experiment also yielded some interesting results pertaining to the perceived place of articulation of the stop consonant, if one was heard. However, we will report these results in a separate paper.

[5] Bailey and Summerfield (in press) showed that the amount of silence needed to hear a stop between a fricative and a vowel decreases as the extent

of the first-formant transition increases. Our [ta] had a much larger first-formant transition than our [tu].

[6] We suspect that the tendency of some subjects to hear stops even at short gap durations was caused by the relatively slow amplitude fall-off (50 msec) of the fricative noise. In pilot studies to Experiment II, we used noises with an abrupt offset, and none of the subjects ever heard stops at the shortest gap durations. Otherwise, the results of these pilot studies were similar to those of Experiment II and therefore are not presented in detail.

[7] In the pilot studies mentioned in Footnote 6, there were two (out of seven) subjects who showed a reduction in the vowel context effect as stop consonants began to be heard. Both subjects, BHR and GE, also participated in the present study, but only one (BHR, one of the authors) continued to show a slight reduction in the context effect at the stop/no-stop boundary.

[8] Interestingly, SL was the only subject in Experiment II showing clear evidence of no context effect at gap durations beyond 40 msec. For her, the context effect could have been entirely due to the transitions.

INFLUENCE OF PRECEDING FRICATIVE ON STOP CONSONANT PERCEPTION

Virginia A. Mann and Bruno H. Repp

Abstract. The effect of a preceding fricative on the perceived place of stop consonant articulation was investigated in three experiments. In Experiment 1, we preceded synthetic syllables from two [tV]-[kV] continua with fricative noises appropriate to [ʃ] or [s] and showed that more velar stops are perceived following [s]. Spectrographic analysis of fricative-stop-vowel utterances suggested an analogous shift in place of stop occlusion (as revealed in formant transitions) following [s]. Experiment 1 also demonstrated a decrease in the magnitude of the perceptual context effect with increased temporal separation of fricative and CV portions, and with introduction of a syllable boundary after the fricative. Experiment 2 suggested that although the effect declines initially with temporal separation, it may persist in reduced form over intervals as long as 375 msec. Experiment 3 revealed that the context effect is categorical in nature: It depends primarily on the phonetic category assigned to the fricative, rather than on the specific spectral properties of the fricative noise. We interpret these results as evidence for a correlation between speech perception and production.

INTRODUCTION

The articulatory gestures appropriate to the phonetic constituents of an utterance are dynamic, overlapping events. As a consequence, the acoustic information for individual phonetic segments is distributed in time and intertwined with information for neighboring segments. Thus, the listener who is recovering the segmental structure of an utterance must not only integrate temporally distributed cues into unitary phonetic percepts; he must also take into account certain dependencies between cues for different phonetic segments. We presume that, in order to accomplish these tasks, he must draw on an implicit knowledge of articulatory dynamics. The fact that such knowledge does, in fact, play a role in speech perception is revealed not only in a

---

Acknowledgment: This research was supported by NICHD Grant HD01994 and BRS Grant RR05596 to Haskins Laboratories, and by NICHD Postdoctoral Fellowship HD05677 to Virginia Mann. Some of our results were reported at the 96th and 97th meetings of the Acoustical Society of America in Honolulu, Hawaii, December 1978, and in Cambridge, Massachusetts, June 1979, respectively; and at the 10th International Congress of Phonetic Sciences in Copenhagen, August 1979. We thank Alvin Liberman for his advice and encouragement at all stages of our research, and Alice Healy and David Isenberg for helpful comments on an earlier version of this paper.

listener's success in dealing with the natural consequences of coarticulation, but also in his perception of synthetic speech in which coarticulatory effects have been deliberately removed or distorted. Just as he is able to achieve perceptual constancy in the face of acoustic variability generated by natural coarticulation, so he may show perceptual variability in the face of (unexpected) local acoustic constancy. Presumably, such perceptual "context effects" reflect the listener's expectations of certain contextual dependencies in the acoustic signal.

The existence of perceptual context effects is well documented: Identical acoustic segments of speech are often interpreted differently in different contexts. Perhaps the best-known example concerns the perception of release bursts preceding a vocalic segment: When a synthetic noise burst with a center frequency of about 1600 Hz is followed by a steady-state vocalic portion appropriate to [i] or [u], it leads to the perception of [pi] or [pu]; however, when the same noise is followed by a vocalic portion perceived as [a], listeners report hearing [ka] (Liberman, Delattre, & Cooper, 1952). Another example concerns an effect of vocalic context on the perception of fricatives. When synthetic fricative noises drawn from a [ʃ]-[s] continuum are followed by various vocalic portions, listeners hear more instances of [s] in the context of rounded vowels such as [u] than in the context of unrounded vowels such as [a] (Kunisaki & Fujisaki, 1977; Mann & Repp, 1979; Whalen, 1979). Both of these perceptual context effects correspond to contextual dependencies in the acoustic signal that are due to coarticulation.

In the present experiments, we investigated a context effect that, to our knowledge, has not been previously described: the influence of a preceding fricative on the perceived place of articulation of a stop consonant. We accidentally discovered this effect in the course of experiments on another context effect, that of vocalic context on fricative perception (Mann & Repp, 1979). In Experiment 1, we assessed this new effect more precisely and also explored how its magnitude is affected by (1) increases in the temporal separation between the fricative noise and the following signal portion, and (2) the introduction of a syllable boundary between fricative and stop. In Experiment 2, we extended our investigation of the effects of temporal separation. In Experiment 3, we investigated whether the effect of the fricative on the stop is due to spectral characteristics of the fricative noise or to the phonetic category to which subjects assign that noise.

EXPERIMENT 1

In this experiment, our goal was to demonstrate that [ʃ] and [s] differentially affect listeners' identification of following stop consonants (drawn from a synthetic [t]-[k] continuum), and to explore some of the conditions that might influence the magnitude of this effect. There were five different experimental conditions. In the first of these (the CV condition), stimuli from [ta]-[ka] and [tu]-[ku] continua were presented for identification. Subjects' responses in this condition provided a baseline measure of stop consonant perception. In the remaining four conditions, the stimuli from the two CV continua were presented in the context of a preceding [ʃ] or [s]. Here we chose to vary orthogonally two factors: the temporal separation of fricative noise (F) and vocalic (CV) portion (75 or 150 msec),[1] and the

75

presence or absence of a vowel preceding the fricative (VFCV vs. FCV). As we will explain below, the second factor essentially concerned the presence or absence of a syllable boundary between fricative and stop.

Method

Subjects. Eleven adults participated as subjects. They included eight paid volunteers with varying experience in listening to synthetic speech, a research assistant, and the two authors. The data of one subject were excluded because of unusually high variability. The data were pooled across the remaining ten subjects.

Stimuli. All stimuli were generated on the OVE-IIIc serial resonance synthesizer at Haskins Laboratories. They included nine CV syllables constituting a [ta]-[ka] continuum (perceived by listeners as ranging from "da" to "ga"), nine syllables constituting a [tu]-[ku] continuum (intended to range from "du" to "gu"), two fricative noises appropriate for [ʃ] and [s], and two steady-state vowels, [a] and [u]. The CV stimuli were periodic throughout, 200 msec in duration, and differed in the onset frequencies of the second and third formants (F2 and F3). For stimuli from the [ta]-[ka] continuum, the duration of the initial formant transitions was 50 msec. Onset frequency of F1 was held constant at 250 Hz; F2 and F3 onset values are listed in Table 1. Steady-state frequencies of the first three formants of [a] were 771, 1233, and 2520 Hz, respectively. For stimuli from the [tu]-[ku] continuum, the durations of F1, F2, and F3 transitions were 30, 70, and 80 msec, respectively. Onset frequency of F1 was held constant at 200 Hz; F2 and F3 onset values are listed in Table 1. Steady-state frequencies of the first three formants of [u] were 250, 800, and 2295 Hz, respectively. In all stimuli, fundamental frequency declined linearly, and amplitude rose linearly during the first 80 msec and remained steady for the remaining 120 msec.

Table 1

Formant transition onset frequencies (Hz) for two CV continua.

	[ta]-[ka]		[tu]-[ku]	
Stimulus No.	F_2	F_3	F_2	F_3
1	1790	2709	1796	2502
2	1770	2576	1744	2379
3	1744	2449	1695	2245
4	1719	2338	1646	2119
5	1695	2197	1600	2000
6	1670	2074	1554	1874
7	1646	1943	1499	1744
8	1623	1821	1456	1622
9	1600	1694	1404	1499

The two fricative noises were the endpoints of a continuum used in a previous study of the [ʃ]-[s] distinction (Mann & Repp, 1979); they were considered to be reasonably appropriate for their respective categories. Each was characterized by two steady-state formants (poles) produced by the fricative circuit of the synthesizer. Formant center frequencies for [ʃ] were 1957 and 3803 Hz; for [s], 3917 and 5077 Hz. Noise duration was 150 msec; amplitude rose during the first 50 msec, remained steady for the next 50 msec and fell during the final 50 msec.

Random stimulus sequences for each of the five conditions were recorded directly from the synthesizer onto magnetic tape. A three-second interval separated individual stimuli, and there were longer pauses between sequences. The *CV condition* employed five sequences of 42 stimuli each. Within each set of 42, half of the stimuli were from the [ta]-[ka] continuum and half were from the [tu]-[ku] continuum; the nine stimuli from each continuum were presented with unequal frequencies according to a 1-2-3-3-3-3-3-2-1 schedule. Stimuli for the two fricative-stop-vowel (FCV) conditions were formed by following the [ʃ] and [s] noises by each of the CV syllables, thus yielding twice as many stimuli as in the CV condition. In the *F(75)CV condition*, F and CV were separated by a 75-msec period of silence; in the *F(150)CV condition*, the silent gap was 150 msec long. The tapes for each of these conditions contained five random sequences of 84 stimuli. Frequency of stimulus presentation followed the same schedule as in the CV condition. Stimuli for the *VF(75)CV and VF(150)CV conditions* were created by immediately preceding the stimuli from the FCV conditions with a steady-state [a] or [u]. The tape for each of these two conditions consisted of four sequences of 168 stimuli each. The first and fourth sequences contained initial [a], the second and third contained initial [u]. The results were pooled across these two initial vowels.

Procedure. Each subject participated in two 90-minute sessions, during which he or she was seated in a quiet room, listening over Telephonics TDH-39 earphones at a comfortable intensity. The CV tape was presented at the beginning of each session, followed by two of the other tapes, chosen to have the same gap duration. Except for the CV tape, order of presentation was counterbalanced across subjects. Each subject gave a total of 15 responses to stimuli drawn from the centers of the CV continua in the CV (heard twice, once in each session) and FCV conditions, 12 responses in the case of the VFCV conditions.

In all conditions, the subjects were asked to identify the consonants heard. Due to the phonology of English, voiceless unaspirated [t] and [k] are transcribed as "d" and "g," respectively, in syllable-initial position, but as "t" and "k," respectively, when preceded by a fricative perceived as belonging to the same syllable. For this reason, the response choices for the stops were "d" and "g" in the two CV conditions, but "t" and "k" in the two FCV conditions. In the VFCV conditions, however, the subjects were again asked to respond with "d" or "g." By these instructions, we effectively induced the subjects to place a syllable boundary between fricative and stop, since syllable-initial [sd], [sg], [ʃd], and [ʃg] clusters are not permitted in English.

Results

Figures 1 and 2 show average percentages of alveolar ("d" or "t") responses as a function of position along each CV continuum. Figure 1 shows the results for the [ta]-[ka] series, Figure 2 those for the [tu]-[ku] series. The response functions obtained in the CV condition are represented by dotted lines. The dotted functions in left and right panels duplicate each other, whereas those in upper and lower panels are independent replications. (Recall that the CV tape was heard twice, to provide separate baselines for conditions differing in gap duration.) The probability of a "t" response was high for the first few stimuli in each continuum and gradually decreased to a minimum for the final few stimuli. Although the two authors, who served as the first subjects, labeled the [tu]-[ku] stimuli very consistently, all other subjects had much difficulty hearing velar stops preceding [u], and gave an overwhelming number of "d" responses. This is evident in Figure 2: The overall percentage of velar stop responses to stimuli intended to sound like [ku] did not exceed 50 percent.[2] Although this was an unwelcome result, the response distributions on the [tu]-[ku] continuum did change considerably when an initial fricative noise was added to the stimuli.

The solid lines represent responses to stimuli in which an [ʃ] preceded the CV portion, and the dashed lines represent responses to stimuli with an initial [s]. Comparison of these functions reveals a large effect of fricative context: Subjects gave many fewer alveolar stop responses in [s] context than in [ʃ] context. This difference was highly significant, both on the [ta]-[ka] continuum, $F(1,9) = 42.87$, $p < .0005$, and on the [tu]-[ku] continuum, $F(1,9) = 36.33$, $p < .0005$. The functions for the two fricative contexts were further distinguished by the extent to which they departed from the CV results. Relative to this baseline, the number of alveolar responses was decreased by the presence of an initial [s] but was not significantly affected by an initial [ʃ].[3]

Inspection of Figures 1 and 2 further reveals the extent to which the magnitude of this context effect was reduced by an increase in the duration of the silent gap and by the presence of a syllable boundary between fricative and stop consonant. The effect of gap duration was significant, $F(1,9)=9.75$, $p < .025$, as was the effect of syllable boundary, $F(1,9)=18.16$, $p < .005$. Although the figures show some variation in the extent of these two different effects on the two stimulus continua, none of the interactions involving stimulus continuum reached significance.

Discussion

Consistent with our expectations, subjects heard more velar stops following [s] than following [ʃ] (or in the absence of any fricative). The difference was remarkably large. How is this effect to be explained? Our theoretical biases lead us to look first towards articulation for possible clues. Our intuitions--and those of our colleagues--as to the articulation of fricative-stop clusters suggest that the place of articulation of velar (and, perhaps, alveolar) stop consonants following [s] may be relatively more forward than that of the same stop consonants following [ʃ]. As yet we have no confirmatory data on articulatory movements. However, we have examined some spectrograms of FCV utterances produced by two male native speakers of

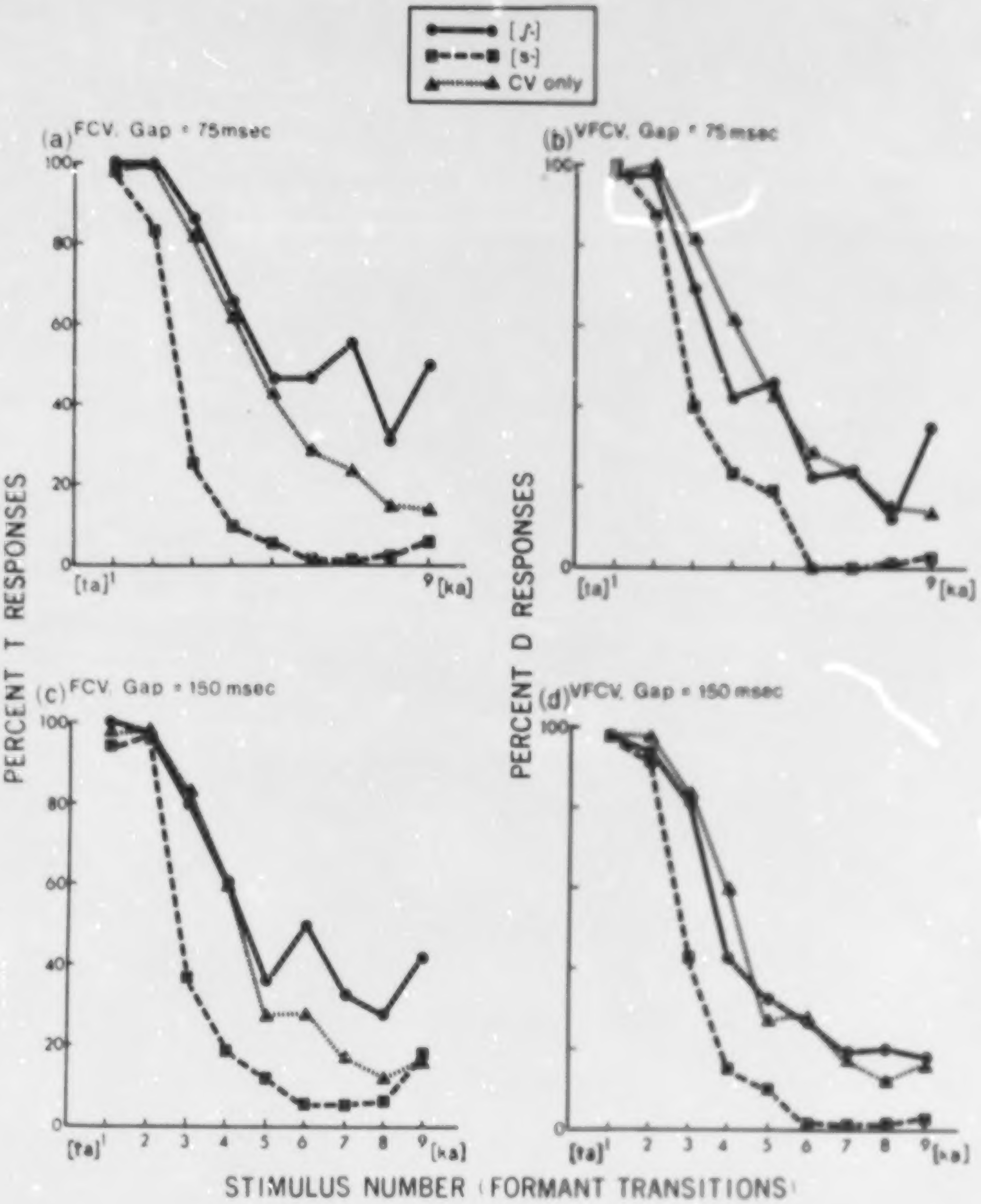


Figure 1. Effects of [ʃ] and [s] on the perceived place of articulation of a following stop consonant, as a function of gap duration and presence vs. absence of a syllable boundary ([ta]-[ka] continuum).

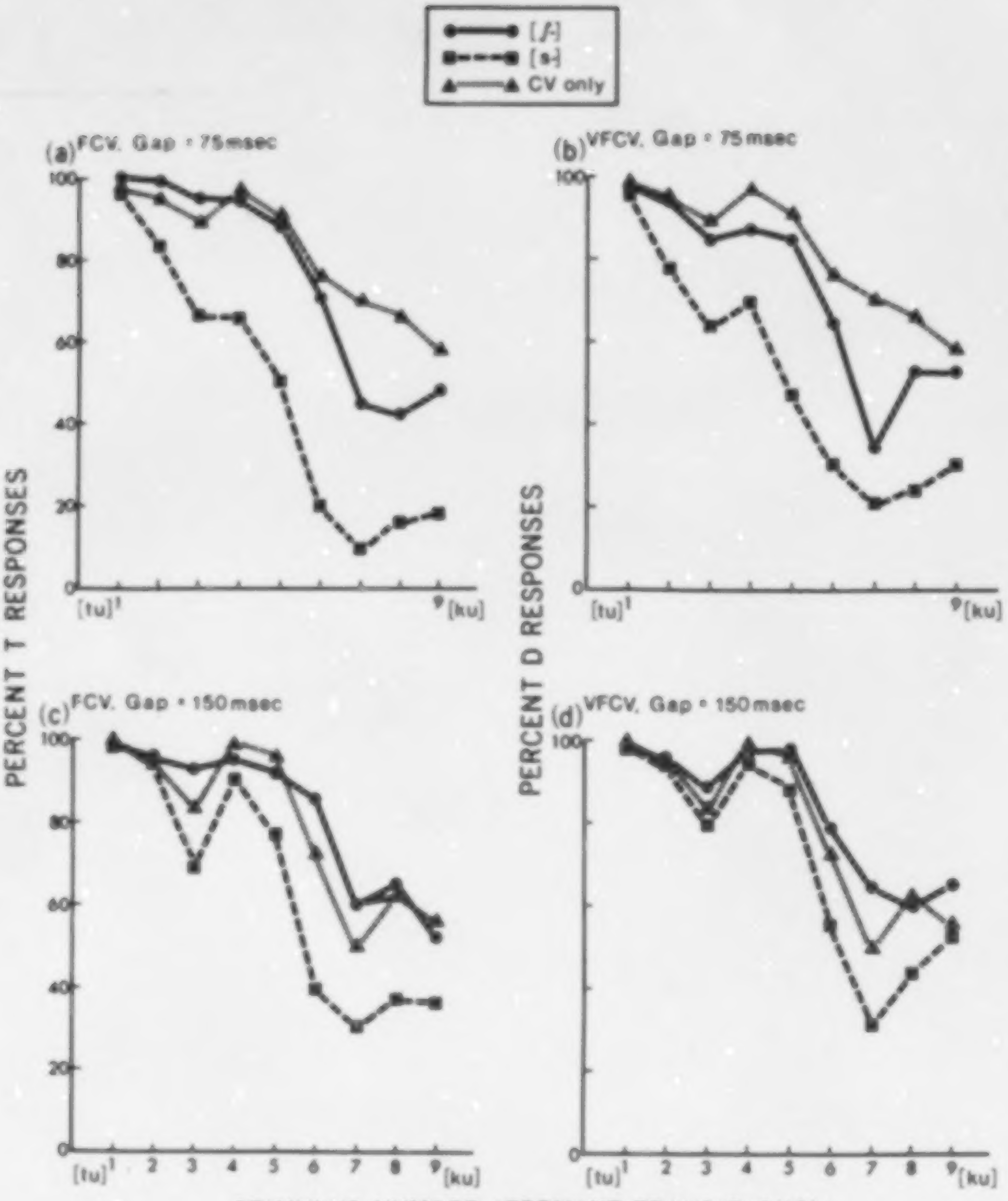


Figure 2. Effects of [ʃ] and [s] on the perceived place of articulation of a following stop consonant, as a function of gap duration and presence vs. absence of a syllable boundary ([tu]-[ku] continuum).

English. They suggest that the onset frequencies of F_2 and F_3 for velar stops are farther apart after [s] than [ʃ], whereas alveolar stops show the opposite. These differences are consonant with our hypothesis of a forward shift in place of tongue-palate contact following [s]. We suggest that listeners possess implicit knowledge of this coarticulatory dependency and compensate for it in perception.

If the basic context effect in perception can be explained on an articulatory basis, then perhaps changes in its magnitude with various experimental factors are to be explained in the same way. For instance, it seems likely that a prolongation of the stop closure interval would weaken the coarticulation of fricative and stop; therefore, a parallel decline in the perceptual context effect may be predicted, in accordance with our findings. A similar prediction may be made in the case of syllable boundaries. We have preliminary spectrographic data showing a dependency of formant transitions on the preceding fricative even when a syllable boundary intervenes, but as yet we cannot tell whether the extent of that dependency is reduced compared to utterances in which fricative and stop are produced as part of the same syllable. However, it would not be far-fetched to expect such a reduction; in fact, one of the cues for syllable (or word) juncture may in fact be reduced coarticulation, if only as a consequence of prolonged closure (cf. Christie, 1974).

Our finding that introduction of a syllable boundary reduces the perceptual context effect not only suggests that an analogous reduction in coarticulation may indeed occur; it also provides an interesting instance of a top-down effect on segmental perception. Note that, when introducing a syllable boundary, we held the local acoustic environment constant; only the instructions to the listeners were varied.[4] The fact that these instructions had perceptual consequences suggests to us that the listeners made use of abstract knowledge about articulatory processes and their changes with phonological structure.

Although we feel that our hypothesis of a perception-production link stands on solid ground, we realize that several alternative explanations may be proposed for the findings of Experiment 1. However, we postpone their discussion until we have described a second experiment whose results, as will be seen shortly, introduce a complicating factor.

EXPERIMENT 2

Our second experiment was designed to explore further the temporal limits of the context effect observed in Experiment 1. We were surprised to find that the fricative still had such a strong influence on stop consonant perception when F and CV were separated by 150 msec. In Experiment 2, we varied the temporal separation between 75 and 375 msec, a value far exceeding normal stop closure durations. We hoped to observe a systematic decline and eventual disappearance of the context effect within that range.

Method

Subjects. Nine adults participated as subjects. They included six paid volunteers and a research assistant, none of whom had participated in

Experiment 1, plus the two authors.

Stimuli. The nine members of the [ta]-[ka] continuum were preceded by either [ʃ] or [s], which was in turn preceded by steady-state [a]. A silent gap of 75, 150, 225, 300, or 375 msec separated the VF and CV portions.

Three sequences of 210 stimuli were recorded, each containing 42 stimuli at each of the five different gap durations. Gap duration varied randomly within each sequence. At each gap duration, half of the stimuli contained initial [s] and half contained initial [ʃ]. Other details were the same as in Experiment 1.

Procedure. Subjects participated in a single 90-minute session in which they listened to the three sequences, paused for a brief rest, and then listened to two of the sequences a second time. Thus, each subject gave a total of 15 responses to stimuli drawn from the center of the [ta]-[ka] continuum. By requiring the subjects to label the stops as "d" or "g," we induced placement of a syllable boundary between fricative and stop. This was done so the long silent intervals would not sound unnatural.

Results

Three of the subjects rarely perceived "g" after [ʃ], although they had no difficulty with stops after [s]. Consequently, they showed exceptionally large context effects at all temporal separations; no systematic decrease was evident, except between the two shortest intervals. The average results of the other six subjects, who showed regular labeling functions, are displayed in Figure 3. They are plotted as the average percentage of "d" responses as a function of silent gap duration. (To simplify presentation, responses have been pooled across the nine members of the [ta]-[ka] continuum.) The solid line represents responses to stimuli that contained [ʃ], the dashed line represents responses to stimuli that contained [s]. As in Experiment 1, there were more "d" responses in the context of [ʃ] than in the context of [s], $F(1,5) = 7.8$, $p < .05$. At the same time, the magnitude of this context dependency decreased significantly with increased temporal separation between fricative and stop, $F(4,20) = 3.3$, $p < .05$. However, Figure 3 shows a sharp decline only between 75 and 150 msec, with little change thereafter. Thus, the results of these six subjects were essentially similar to those of the three subjects who had difficulty hearing "g" after [ʃ]. One puzzling feature of these data is that, although Experiment 1 showed that the context effect was primarily due to [s], it was the effect of [ʃ] that changed with temporal separation in the present study. The results do suggest, however, that some context effect persists over temporal separations of more than 375 msec.

Discussion

The temporal persistence of the fricative-stop context effect certainly came as a surprise. Not only were the longest intervals used here far beyond the durations of natural stop closures, but the listeners were also led to assume a syllable boundary between fricative and stop, which should have reduced the context effect to begin with. The data suggest, however, that there may be two components to the context effect, possibly with different underlying causes. One component rapidly declines with temporal separation

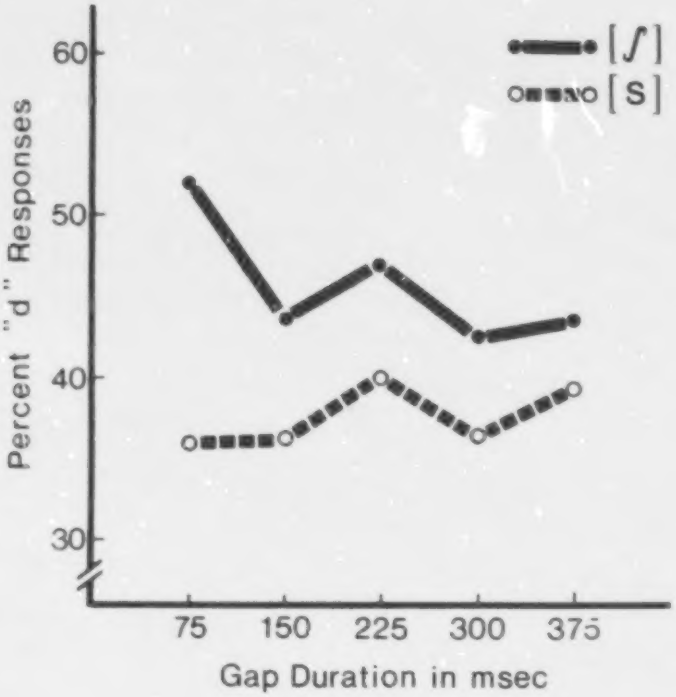


Figure 3. Effects of [ʃ] and [s] on perceived place of stop consonant articulation, as a function of gap duration.

and disappears around 150 msec; the other component changes little with temporal separation and persists over intervals beyond 375 msec. This interpretation--which we will adopt as a working hypothesis--complicates the theoretical explanation of the context effect.

To the extent that perceptual context effects mirror coarticulation, they would be expected primarily when there is indeed a contextual constraint on articulation, i.e., when a speaker has no choice but to coarticulate. Such a constraint probably exists between fricative and stop production as long as the stop closure interval is sufficiently short. The first, rapidly decaying component of the context effect may reflect the listeners' compensation for that constraint. However, at closure intervals of 300 msec or longer, a speaker certainly has the option of articulating the stop release quite independently of the preceding fricative. We note, though, that he also has the alternative option of establishing a fricative-dependent tongue-palate contact right at the offset of frication and to maintain that position until the release, no matter how long the closure. Which of these alternative strategies is more commonly chosen by speakers, we do not know. It would seem surprising to us if perception reflected an articulatory phenomenon that is not obligatory; however, this possibility cannot be ruled out with certainty. Thus, the hypothetical second, long-lasting component of the context effect may have an articulatory basis, too, although this seems less likely.[5] The time has come to consider alternative explanations.

Three hypotheses need to be considered, in addition to the postulated perception-production relation discussed up to now. The first of these is a response bias account. According to this hypothesis, listeners are, for some reason, biased towards [sk] or [s-g] responses and against [st] or [s-d] responses. (Note that [ʃt] and [ʃk], which do not occur in English in syllable-initial position and therefore may be suspect, are irrelevant to this discussion, since [ʃ]--in Experiment 1, at least--did not exert a significant context effect.) Obviously, such a bias is a candidate explanation of the long-lasting component of the context effect, and it might also be compatible with an effect of syllable boundary; however, it cannot account for the effect of gap duration (the hypothetical short-lived component). One prediction that should hold for any type of response bias is that it should depend on the perceived category of the fricative, not on the precise characteristics of its acoustic cues, such as fricative noise spectrum. This prediction was examined in Experiment 3.

A second alternative explanation of the present context effect is that the [s] noise imposed a psychoacoustic transformation, e.g., through auditory contrast, on the formant transitions of the CV syllables, such that the transitions appeared more appropriate for a velar place of articulation. For example, a preceding high-frequency noise might lower the perceived onset frequency of the third-formant transition, thus increasing the compactness of the vocalic onset spectrum--a cue for velar place of articulation (Stevens & Blumstein, 1978). Such a contrast effect would be consistent with the fact that increased temporal separation of fricative and stop reduced the magnitude of the context effect; thus, it could account for its short-lived component. However, it would be incompatible with the long-lasting component and with the effect of a syllable boundary. In contrast to the response bias explanation, the auditory contrast hypothesis predicts variations in the magnitude of the context effect with changes in fricative noise spectrum, regardless of perceived category. This prediction was tested in Experiment 3.

Yet another hypothesis needs to be considered: that the offset spectrum of the fricative noise provided a cue to place of articulation of the stop, which was perceptually integrated with the formant transition cues. A number of studies have demonstrated that fricative noise contains cues to the place of a following stop occlusion (Uldall, 1964; Malecot & Chermak, 1966; Schwartz, 1967; Bailey & Summerfield, in press). To explain the present results, the steady-state [s]-noise must have provided a cue for velar, rather than alveolar, place of occlusion. On the contrary, studies by Uldall (1964) and Schwartz (1967) suggest that steady-state [s]-noise favors [t] percepts. Malecot and Chermak (1966), on the other hand, report data from a systematic study with synthetic speech, which lead to the opposite conclusion. When listeners had to identify syllable-final stop consonants from frequency changes in [s]-noise alone, "k" responses were more frequent than "t" responses following a steady-state noise. For more reliable [t] percepts, an upward transition was required in the fricative noise. Malecot and Chermak also cited parallel observations in spectrograms of natural speech. Thus, what we might call the "noise-offset-cue hypothesis" cannot be rejected at this point as a possible account for the fricative-stop context effect. (In fact, if the hypothesis were true, we would not be dealing with a true context effect at all.) However, the hypothesis fails to account for any context effects beyond temporal separations of 200 msec, since this seems to be the upper limit of temporal cue integration for stop place of articulation (Repp, 1978; Repp, Liberman, Eccardt, & Pesetsky, 1978). In addition, the hypothesis predicts, as does the auditory contrast hypothesis, that listeners should be quite sensitive to the spectral characteristics of the fricative noise. This prediction was tested in Experiment 3.

EXPERIMENT 3

Experiment 3 was designed to provide information crucial to deciding between some of the alternative hypotheses discussed above. The question of interest was whether the effect of the fricative on the following stop was primarily a function of the fricative's noise spectrum or of its perceived phonetic category. To that end, we examined the effects of several fricative noises ambiguous between [ʃ] and [s] on the [ta]-[ka] distinction. A finding that the number of velar responses given in the context of such stimuli is a continuous function of their respective noise spectra would be consistent with spectral contrast as the basis of the effects observed in Experiment 1. It would also be consistent with an explanation couched in terms of fricative offset spectrum being integrated with the transitional cues to stop consonant place of articulation. However, if assigned phonetic category should prove the major determinant, rather than noise spectrum *per se*, a response bias explanation would be favored. The articulatory hypothesis--which, of course, is our choice--could be twisted to accommodate either outcome; however, it actually seems to fit better with the result that was obtained, as we will argue below. Primarily, though, Experiment 3 was meant to rule out some of the other competing hypotheses.

Method

Subjects. Ten adults served as subjects. They included eight paid volunteers, five of whom had also participated in Experiment 2, and both authors.

Stimuli. The 25 test stimuli used in this experiment were formed by pairing each of 5 fricative noises with each of 5 CV stimuli, separated by a constant 75-msec period of silence. The fricative noises included the two stimuli employed in Experiments 1 and 2 and three noises ambiguous between [ʃ] and [s]. The ambiguity of these three additional noise stimuli was known from an earlier study of the [ʃ]-[s] distinction (Mann & Repp, 1979); they were stimuli 4, 5, and 6 from a nine-member noise continuum whose endpoints were the unambiguous [ʃ] and [s]. Center frequencies of the two formants of each noise are listed in Table 2. Noise duration was 150 msec. The CV stimuli were drawn from the [ta]-[ka] continuum employed in Experiments 1 and 2. They included the two endpoint stimuli and the three stimuli (4, 5, and 6) most ambiguous between [ta] and [ka] (cf. Figure 1).

Table 2

Pole center frequencies (Hz) of fricative noises in Exp. 3.

Stimulus No.	P_1	P_2
1	1957	3803
4	2690	4269
5	2933	4394
6	3199	4655
9	3917	5077

Five randomized sequences of 55 stimuli were recorded. Within each sequence, stimuli that contained an unambiguous F and an unambiguous CV were presented once, stimuli that contained one ambiguous component were presented twice, and those that contained two ambiguous components were presented three times.

Procedure. Each subject participated in a single 1-hour session, in which the test tape was presented twice. Thus, each subject gave a total of 30 responses to stimuli in which both the F and CV components were (more or less) ambiguous. The task was to identify the fricative-stop cluster as "st," "sk," "sht," or "shk."

Results

The results are plotted in Figure 4 as a function of fricative noise spectrum. The probability of "s" responses (solid line) was almost zero for stimuli containing the most [ʃ]-like noise; it gradually increased among mid-

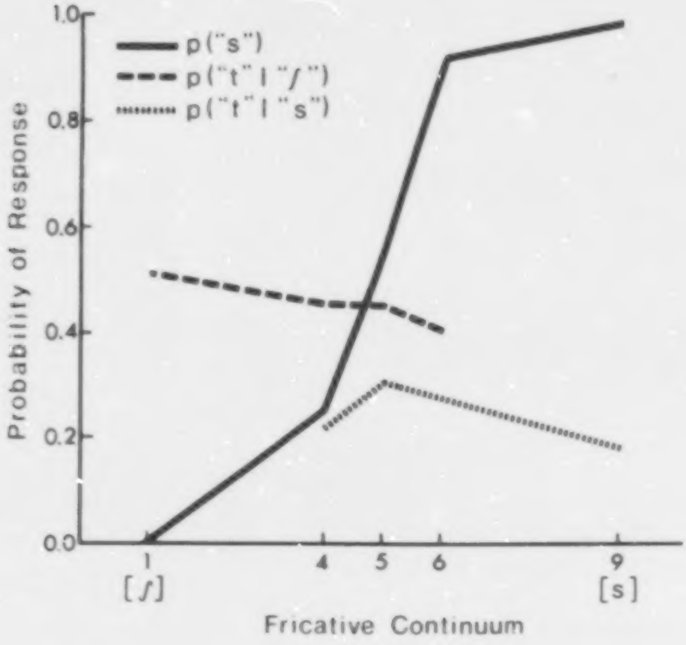


Figure 4. Effect of perceived fricative category on the place of articulation assigned to a following stop.

87

range noises, and approached 1.0 for the most [s]-like noise. The dashed and dotted lines, respectively, represent the probability of "t" responses conditional on whether a given fricative was labeled as "sh" or as "s." To simplify presentation, these conditional probabilities have been averaged over the five CV stimuli. The mean probability of "t" responses contingent on an "sh" response was 0.46, and remained relatively stable across changes in fricative noise spectrum. The mean probability of "t" responses contingent on an "s" response was considerably smaller, 0.25, and likewise showed little change with fricative noise spectrum. An analysis of variance computed on responses to stimuli containing the three most ambiguous fricative noises revealed a significant effect of perceived fricative category, $F(1,9) = 9.0$, $p < .025$, but no significant effect of noise spectrum within categories, $F(2,18) = 0.3$. Even when the unambiguous fricative noises were incorporated into the analysis, there was no significant effect of spectrum. Of the ten subjects tested, only two deviated from this pattern. One displayed a relatively random pattern of responses. The other one was the second author; for him, the major determinant of "t" responses appeared to be fricative noise spectrum, rather than phonetic category. This difference may have been a consequence of his extended experience with synthetic speech.

Discussion

The results of Experiment 3 are surprisingly clear in showing a strong effect of perceived fricative category but little effect of fricative noise spectrum on perception of the following stop consonant.[6] Clearly, this finding constrains the possible explanations of the context effect. It is particularly damaging to the two explanations based on auditory properties of the fricative noise, for it suggests that neither auditory contrast nor noise-offset cues to place of stop occlusion are involved to any significant degree. The fact that a relatively short temporal separation of F and CV was used (75 msec) only strengthens that conclusion, since both types of auditory interactions would be expected only at short temporal separations. On the other hand, the results of Experiment 3 do fit the response bias hypothesis.

What could be the origin of such a bias? The only possibility that occurs to us at this point--in our opinion, an unlikely one to begin with--is that the bias arises from unequal frequencies of [s]+stop sequences in the language. Using a standard word-frequency count (Kučera & Francis, 1967), we added up the frequencies of all words beginning with these consonant clusters. We found that not only is [ska] less than one-fifth as frequent as [sta], but [sk] is, in general, less than one-third as frequent as [st]. (The situation is reversed for [stu] and [sku], however.) If frequency of occurrence had influenced our subjects' responses, then fewer velar stops should have been reported in the context of [s]. Clearly, our finding that [s] leads to an increased number of velar stop percepts does not favor such an account. It is conceivable, however, that a response bias still exists, due to some as yet unknown cause. However, as we have no clue as to the possible origin of such a bias, the hypothesis may be essentially vacuous. We are left with the articulatory account as the most plausible alternative.

The reader will recall our suggestion that the increased number of velar responses in the context of [s] arises from listeners' implicit knowledge of the fact that the transitions for velar and alveolar stops vary with the nature of the preceding fricative, reflecting a forward shift in the place of tongue-palate contact following [s]. However, in order to make this hypothesis consistent with the categorical effect observed in Experiment 3, it is necessary to assume that listeners assigned the synthetic fricative noises to a canonical or idealized place of articulation before utilizing their knowledge of fricative-induced shifts in the place of stop consonant articulation. The listener's abstract knowledge about coarticulatory shifts almost certainly is represented in similarly discrete terms, i.e., in modifications of the canonical forms; we find it hard to imagine that it is based on storage of all possible coarticulatory patterns in analog form (although precisely this possibility has recently been suggested, at least within the domain of computer speech recognition--Klatt, 1979). Thus, the finding of Experiment 3 that the effect of a fricative on a following stop is categorical in nature seems to fit quite well with an account based on perception-production relationships.

Experiment 2 suggested that there may be two components to the context effect. The results of Experiment 3, on the other hand, did not differentiate between them. If there are two components at all, they both seem to represent a categorical effect. But, with some imagination, the articulatory hypothesis may be extended to accommodate even two components: One component may represent the listener's compensation for obligatory constraints in production, whereas the other, long-lived component may reflect his knowledge of optional articulatory strategies, or, alternatively, a perceptual idealization of articulatory principles beyond the region where they normally apply. Thus, the "bias" suggested by the context effects at long temporal separations (Experiment 2) may be based on articulatory principles, after all.

In conclusion, the present findings constitute yet another--and, in our eyes, a particularly convincing--example of a perception-production relationship. They support the view that phonetic perception requires knowledge about articulatory processes and articulatory dependencies--knowledge that, presumably, only human listeners possess. Speech perception involves tracking the behavior of a dynamic sound-generating mechanism--the vocal tract--and adjusting for the effects of its physical and functional constraints on the speech signal, as revealed in listeners' responses when those effects are artificially removed or distorted. Therefore, it is likely that speech perception cannot be understood without considering the articulatory origin of the signal that is perceived.

REFERENCES

Bailey, P. J., & Summerfield, A. Q. Some observations on the perception of [s]+stop clusters. Journal of Experimental Psychology: Human Perception and Performance, in press.

Christie, W. M., Jr. Some cues for syllable juncture perception in English. Journal of the Acoustical Society of America, 1974, 55, 819-821.

Dorman, M. F., Raphael, L. J., & Liberman, A. M. Some experiments on the sound of silence in phonetic perception. Journal of the Acoustical Society of America, 1979, 65, 1518-1532.

Dorman, M. F., Studdert-Kennedy, M., & Raphael, L. J. Stop-consonant recognition: Release bursts and formant transitions as functionally equivalent, context-dependent cues. *Perception & Psychophysics*, 1977, *22*, 109-122.

Klatt, D. H. Speech perception: A model of acoustic-phonetic analysis and lexical access. In R. Cole (Ed.), *Perception and production of fluent speech*. Hillsdale, N.J: Erlbaum, 1979.

Kucera, H., & Francis, W. N. *Computational analysis of present-day American English*. Providence, R.I.: Brown University Press, 1967.

Kunisaki, O., & Fujisaki, H. On the influence of context upon perception of voiceless fricative consonants. *Annual Bulletin of the Research Institute of Logopedics and Phoniatrics*, 1977, *11*, 85-91.

Liberman, A. M., Delattre, P. C., & Cooper, F. C. The role of selected stimulus variables in the perception of the unvoiced stop consonants. *American Journal of Psychology*, 1952, *65*, 497-516.

Malecot, A., & Chermak, A. Place cues for /ptk/ in lower cut-off frequency shifts of contiguous /s/. *Language and Speech*, 1966, *9*, 162-169.

Mann, V. A., & Repp, B. H. Effect of vocalic context on the [ʃ]-[s] distinction: I. Temporal factors. *Haskins Laboratories Status Report on Speech Research*, 1979, *SR-59/60*, this issue.

Repp, B. H. Perceptual integration and differentiation of spectral cues for intervocalic stop consonants. *Perception & Psychophysics*, 1978, *24*, 471-485.

Repp, B. H., Liberman, A. M., Eccardt, T., & Pesetsky, D. Perceptual integration of acoustic cues for stop, fricative, and affricate manner. *Journal of Experimental Psychology: Human Perception and Performance*, 1978, *4*, 621-637.

Schwartz, M. F. Transitions in American English /s/ as cues to the identity of adjacent stop consonants. *Journal of the Acoustical Society of America*, 1967, *42*, 897-899.

Stevens, K. N., & Blumstein, S. E. Invariant cues for place of articulation in stop consonants. *Journal of the Acoustical Society of America*, 1978, *64*, 1358-1368.

Uldall, E. Transitions in fricative noise. *Language and Speech*, 1964, *7*, 13-14.

Whalen, D. H. Effects of vocalic formant transitions and vowel quality on the English /s/-/s/ boundary. *Haskins Laboratories Status Report on Speech Research*, 1979, *SR-59/60*, this issue.

FOOTNOTES

[1] It was necessary that these stimulus portions be separated by at least 50 msec of silence in order to assure reliable perception of a stop consonant (see, e.g., Bailey & Summerfield, in press).

[2] This result was not entirely unexpected. Several of our colleagues have told us of their difficulty in synthesizing [ku] without bursts (cf. Dorman, Studdert-Kennedy, & Raphael, 1977, concerning the importance of the burst in natural tokens of [ku]).

[3] As can be seen in Figures 1 and 2, an initial [ʃ] increased the number of alveolar responses on the [ta]-[ka] continuum, relative to the baseline, but had the opposite effect on the [tu]-[ku] continuum. This was particularly true for the FCV-75 condition, $F(1,9) = 5.96$, $p < .05$, for that interaction.

[4] It seems exceedingly unlikely that, in VFCV stimuli, the initial vowel *per se* would have reduced the effect of the fricative on the stop. However, the fact remains that presence of an initial vowel was confounded with instructions to place a syllable boundary between fricative and stop. We hope to rectify this situation in a future experiment in which we will vary instructions only, to the effect that VFCV stimuli are perceived as either VF-CV or V-FCV.

[5] We are reminded, in this connection, of a possibly relevant finding by Dorman, Raphael, and Liberman (1979): After demonstrating that the insertion of a sufficient amount of silence between the fricative noise and vocalic portions of [slɪt] yields [splɪt], they found that as much as 650 msec of silence was needed to change perception back to [s-lɪt]. Since this interval may be taken as an estimate of the separation necessary to perceive the two signal portions as unrelated to each other, perhaps a similarly long interval would be required to completely eliminate the present context effect.

[6] We have replicated these findings in another study whose results will be reported elsewhere since its main purpose was different (Mann & Repp, 1979). We again found a large effect of perceived fricative category, whereas effects of noise spectrum were inconsistent at best.

ELECTROMYOGRAPHIC STUDY OF THE JAW MUSCLES DURING SPEECH*

Betty Tuller,+ Katherine S. Harris++ and Bob Gross+++

Abstract. An investigation of the role of the mandibular muscles during the production of speech is reported here. Electromyographic (EMG) recordings from superficial and deep masseter, anterior and posterior temporalis, medial pterygoid, superior and inferior lateral pterygoid, and the anterior belly of the digastric muscles were obtained for four speakers of American English. For one of the speakers, mandibular movement was monitored simultaneously with the EMG recordings using a modified thyroumbrometer (Ewan & Krones, 1974). The subjects read lists of nonsense words containing the vowels /a/, /i/ and /u/ in VCV combination with the consonants /p/, /t/, /k/ and /f/. Results indicate that the traditional classification of masseter, temporalis and medial pterygoid as jaw elevators, and lateral pterygoid and anterior belly of digastric as jaw depressors is not adequate for describing control of the jaw in speech.

One focus of current speech research has been the spatial and temporal coordination among articulators, particularly compensation to perturbed (Folkins & Abbs, 1975) or restricted movement of a given articulator (Lindblom, Lubker, & Gay, 1979; Gay & Turvey, 1979) or to a change in the shape of an articulator (Hamlet & Stone, 1976, 1978). The jaw lends itself to studies of this kind because it is intimately involved with other articulators, particularly the lips and tongue, yet is accessible to experimental manipulation.

In the study of interarticulator coordination, analysis of simultaneous movement and electromyographic information should result in a more complete understanding of articulatory control than that allowed by either source alone. However, attempts to execute such experiments are hampered by the ambiguity of existing evidence as to the muscles responsible for movement of the mandible during speech. The assumption that muscles that effect mandibular movement during chewing have similar speech functions may be incorrect.

---

*Accepted for publication in Journal of Phonetics (in press).

+Also University of Connecticut, Storrs.

++Also The Graduate School, City University of New York.

+++Presently at The Medical Center, Louisiana State University.

Acknowledgment: This work was supported by NINCDS grants NS-13617, NS-13870 and BRS grant RR-05596 to Haskins Laboratories. We wish to thank T. Baer, T. Gay, W. Ewan, K. Kirchmeier and S. Josell for making equipment and facilities available to us and ensuring that all ran smoothly, and F. Bell-Berti for comments on an earlier version of this paper.

The chewing cycle must minimally be viewed as consisting of lowering, raising, and occlusal phases. The jaw muscles undergo isotonic contraction during lowering and raising of the mandible, but during the final part of the raising phase, muscular contraction gradually changes from isotonic to isometric (Dubner, Sessle, & Storey, 1978). Isometric contraction during occlusion allows for the development of tension necessary to crush and grind food. In contrast, during speech gestures, the jaw muscles rarely, if ever, undergo isometric contraction with the concomitant development of tension. Moreover, there is evidence that the lowering and raising phases of the masticatory cycle have a different mandibular trajectory from jaw lowering and raising during speech (Gibbs & Messerman, 1972). For example, the lateral excursions evident in chewing are largely absent from speech movements, and the vertical lowering of the jaw in chewing is typically 2-4 times greater than the vertical lowering of the jaw in speech.

The latter difference between speech and chewing gestures is particularly intriguing in light of the complex nature of the temporomandibular joint (Sarnat, 1964; Sicher & DuBrul, 1970). The temporomandibular joint has two compartments, an upper one in which the condyles undergo translation, and a lower one in which the condyles rotate on a hinge axis so that lowering and raising the mandible is not effected by a simple hinge movement of the mandibular condyles. Lowering the mandible combines initial forward translation of the condyles with subsequent rotation. In other words, in lowering the jaw the mandibular condyles move forward and rotate downward. Raising the jaw is a reversal of the lowering gesture such that the condyles rotate upward, then translate backward. Insofar as both the magnitude of jaw lowering and the degree of condylar rotation are smaller for speaking than chewing, the two events imply different relationships between the muscular determinants of condylar translation and rotation. Muscles active during mandibular gestures that are not normal components of speech (e.g., clenches, extreme retrusions and hinge openings) cannot be assumed to function during speech.

Speech and non speech may also differ with respect to the average speed of jaw movement; jaw movements that are speech gestures are typically faster than non-speech jaw movements. The speed of movement required of an articulator is functionally related to the contraction time of the motor units within the relevant muscles. Contraction time (the time from initiation of the twitch to peak tension) has been measured in at least three jaw-raising muscles--medial pteryoid (MacNeilage, Sussman, Westbury, & Powers, 1979), masseter and temporalis (Yemm, 1977). Mean contraction time of medial pterygoid is approximately one-half the mean contraction time of masseter and temporalis. If the different contraction times reflect different speeds of movement required of the muscles, then speech and non-speech gestures may be effected by different sets of muscles. Specifically, the longer contraction times of masseter and temporalis would be suited to non-speech movements, whereas the shorter contraction time of medial pterygoid would be more suited to speech gestures.

The muscles believed to effect non-masticatory mandibular movements (other than speech) are as follows:

MUSCLE DESCRIPTIONS

Masseter is a thick, powerful muscle that runs from the zygomatic bone to the mandible. It has a superficial portion in which the fibers run down and back to the angle of the mandible, and a deep portion in which the fibers are more nearly vertical. It is generally accepted that masseter elevates and clenches the jaw (Ahlgren, 1966; Møller, 1966, 1974; Woelfel, Hickey, Stacy, & Rinear, 1960). Superficial masseter also acts to protrude the jaw, whereas deep masseter acts to retrude the jaw.

Temporalis is a large, fan-shaped muscle that runs from the lateral surface of the cranium to the coronoid process and ramus of the mandible. The fibers of the anterior portion run almost vertically; hence, their line of pull acts to elevate the mandible. The posterior portion of temporalis runs horizontally forward to the anterior edge of the root of the zygoma. The fibers then bend downward and attach to the mandibular notch. Posterior temporalis elevates and retrudes the jaw and moves it laterally.

Medial (internal) pterygoid runs parallel to the masseter but is deep to the mandible. Together, masseter (particularly the superficial portion) and medial pterygoid form a sling around the angle of the mandible, pulling upward and forward, providing a mechanism for powerful elevation and clenching of the jaw.

Lateral (external) pterygoid is composed of two partially discrete portions, both of which run from the outer surface of the lateral pterygoid plate to the neck of the mandible. The fibers of the superior portion run horizontally, whereas the fibers of the inferior portion run in a forward and upward direction. Lateral pterygoid appears to act during mandibular protrusion. Electromyographic activity has also been recorded in this muscle during both depression and elevation of the mandible. Hickey, Stacy, and Rinear (1957), Møller (1966), and Woelfel, Hickey, Stacy, and Rinear (1960) found lateral pterygoid to be active during jaw lowering, whereas Carlsöo (1956), Hickey et al. (1957), Møller (1966), and Griffin and Munro (1969) recorded lateral pterygoid activity during jaw elevation. Recent evidence has suggested that the inferior and superior heads of lateral pterygoid function independently (Grant, 1973; McNamara, 1973); activity was evident in the superior head during jaw raising but not lowering or protrusion, whereas the inferior head was active during jaw lowering and protrusion.

The anterior belly of digastric (ABD) runs from the deep surface of the body of the mandible to the hyoid bone. It is generally referred to as a depressor of the mandible, although it may also function to stabilize the hyoid bone. It is maximally active when lowering the jaw against resistance (Griffin & Malor, 1974).

Mandibular muscles shown to be involved in chewing and large temporomandibular movements have been, in the past, studied in experiments involving speech gestures. For example, Sussman, MacNeilage, and Hanson (1973) used electromyography (EMG) to study labial-mandibular control and coordination in speech. They recorded EMG activity from masseter, medial pterygoid, and anterior belly of digastric muscles but found only the digastric to be active for speech gestures. Neither masseter nor medial pterygoid could be consistently related to jaw movement. Folkins and Abbs (1975) also monitored EMG activity in masseter and medial pterygoid, as well as in anterior temporalis. In contrast with Sussman et al. (1973), Folkins and Abbs found that medial pterygoid was indeed consistently related to jaw movement in speech, although masseter and anterior temporalis were not. More recently, Folkins, Zimmerman, and Cooper (Note 1) did find low levels of speech-related activity in masseter and temporalis. Medial and lateral pterygoid were active at higher levels although lateral pterygoid activity was not related to jaw movement in a consistent fashion.

Without more precise information as to which muscles function to control the mandible during normal speech gestures, it is impossible to obtain an accurate account of the coordination of the jaw with other articulators, or articulatory compensation to restricted or perturbed movement. Accordingly, the experiment reported here examined the functional role of certain mandibular muscles during the production of a small inventory of speech gestures. As an example of how these data may be used, we also examined the effect of different phonetic environments on muscle activity for a given speech gesture; these data will be reported separately.

METHOD

EMG recordings were collected using bipolar hooked-wire electrodes of the type described by Hirose (1971). During insertion of the electrodes the subject was in a slightly reclined position and breathed nitrous oxide to reduce discomfort. Detailed descriptions of electrode placement and insertion techniques may be found in Ahlgren (1966) and Gross and Lipke (Note 2). Verification of electrode placement used those non-speech maneuvers for which each muscles role is well established (Ahlgren, 1966; Carlsöo, 1952, 1956; Møller, 1966, 1974; Moyers, 1950).

Masseter: Activity from the superficial and deep portions was recorded separately. Placement of the electrode in the superficial portion was verified by its activity during protrusion of the mandible and clenching. Placement in the deep portion was verified by its activity during clenching of the mandible.

Temporalis: EMG activity was recorded separately from anterior and posterior temporalis. Electrode placement in the anterior portion was verified by activity in the muscle during elevation, but not retrusion, of the mandible. Electrode placement in the posterior portion was verified by its activity during retrusion of the mandible.

Medial (internal) pterygoid: Placement was verified by activity during elevation and clenching of the mandible.

Lateral (external) pterygoid: Activity from the two heads was recorded separately. Electrode placement in the superior head was verified by its strong activity during clenching but not protrusion of the mandible. Placement of the electrode in the inferior head was verified by its activity during protrusion of the mandible.

The anterior belly of the digastric (ABD): Verification of electrode placement was achieved by its activity during large-excursion jaw lowering, particularly against resistance.

The EMG potentials were recorded onto magnetic tape, rectified, subsequently soft-ware integrated with a time constant of 35 ms., and averaged using the Haskins Laboratories EMG system described by Kewley-Port (1973, 1974).

The subjects were three adult females and one adult male. Three of the four subjects were naive as to the purpose of the experiment. Subjects BE and CC have Class I occlusions, following Angle's classification of the forms of occlusion (Kerr, Ash, & Millard, 1978), showing a normal relationship between maxillary and mandibular dentition. Subject BT has a Class II occlusion with the mandible in a posterior relationship to the maxilla. Subject VR has a Class III occlusion, in which the mandible is protruded relative to its normal relationship with the maxilla.

Jaw displacement in the vertical and horizontal dimensions was measured for Subject CC, simultaneously with the recording of EMG potentials using a modified version of the thyroumbrometer (Ewan & Krones, 1974). This device consists of an array of photocells and a PDP 11/34 computer. An inflexible pointer that had been custom-made for the subject was extended from her lower teeth. A dc light source cast the shadow of the pointer onto the photocells of the thyroumbrometer. The vertical and horizontal position of the jaw was computed from the photocell voltages. The computer output voltage was a staircase function, each step change indicating a .5 mm change in vertical jaw position; horizontal jaw position could not be measured to the same degree of accuracy. The jaw displacement signal and the EMG potentials were recorded simultaneously onto separate channels of an FM data recorder.

The speech utterances were four-syllable nonsense words of the form /əkV_1CV_2pə/. In all cases, either V_1 or V_2 was /a/, whereas the other vowel varied among the set /a,i,u/. The consonant (C) was either /p/,/t/,/k/ or /f/. The 20 utterance types were randomly ordered and repeated six times at a comfortable speaking rate. Subjects were instructed to produce the second and third syllables of the utterance with equal stress, with the first and last syllables unstressed. The end of periodicity in the acoustic signal of the first vowel (V1) was the point chosen for aligning tokens for averaging, and is represented by the zero point on the abscissa in the figures.

One year and ten months after the original EMG recording session, Subject VR repeated the experiment with electrode insertions into the anterior belly of the digastric, medial pterygoid and both heads of lateral pterygoid. The original 20 utterance types were randomly ordered and repeated 12 times at a comfortable speaking rate. All other instructions were identical to the first recording session and EMG data from both sessions were processed in identical fashion.

RESULTS

The patterns of EMG activity will be presented separately for non-speech maneuvers, speech gestures specific to phonetic segments, and coarticulation.

Within these divisions, the pattern of activity of each muscle will be discussed. The magnitude of the speech-related activity was assessed as a percent value of the maximum level of activity achieved during non-speech gestures.

Non-speech Maneuvers

Deep and superficial masseter functioned independently for all subjects. Deep masseter was always active during retrusion and clenching of the mandible. Subject CC used deep masseter to elevate the jaw from extremely low positions. In contrast, superficial masseter was active during protrusion and clenching of the mandible for all subjects, and during elevation of the mandible for three subjects (BT, CC, VR).

Anterior temporalis was active for all subjects during mandibular elevation and clenching. Posterior temporalis was active for all subjects during mandibular retrusion and clenching.

Medial pterygoid was consistently active for clenching, elevation and protrusion of the mandible. For two subjects (BE, CC) medial pterygoid also functioned during retrusion of the jaw.

Superior head of lateral pterygoid was active during jaw elevation and clenches (all subjects).

Inferior head of lateral pterygoid was active during mandibular protrusion and depression for all subjects. For two subjects (BE, VR) this muscle showed low levels of activity during mandibular retrusion.

Anterior belly of the digastric acted during large-excursion lowering and retrusion of the jaw for all subjects, and during protrusion of the jaw for subjects BE, CC, and VR.

These results are in general agreement with the investigations of activity in mandibular muscles during chewing and large temporomandibular movements (cf. Gross & Lipke, Note 2), and imply that electrode placements did not change significantly between the time of verification and the actual experimental maneuvers. In a later section, the levels of activity recorded during the non-speech gestures will be compared with the levels of activity recorded during speech.

Jaw Displacement

Measurements of mandibular displacement during speech were obtained for Subject CC; no non-speech maneuvers were performed. As noted above, the accuracy of the system for monitoring jaw movement in the horizontal plane is significantly less than the accuracy of monitoring vertical movement. Consequently, the measurements were sufficient to detect systematic movement patterns in the vertical, but not the horizontal, dimension.

In vowel production, vertical mandibular positions were lowest for /a/ and highest for /u/. During the onset of consonant constriction the mandible was highest for /t/ and /f/, slightly lower for /p/ and lowest for /k/.

Electromyographic differences were examined to determine whether they reflect the differences in vertical jaw displacement. For this subject, the level of activity in the most consistently active jaw depressor (inferior head of lateral pterygoid) was directly related to the amount of mandibular lowering required for production of the vowel (Figure 1). Activity in the jaw elevator (medial pterygoid) did not change systematically as a function of mandibular height required for a given consonant. Activity was greatest when moving from the lowest jaw position, that is from the vowel /a/, to the following consonant, regardless of the identity of the consonant. EMG recordings from all subjects were examined to determine which muscles were responsible for the jaw movements during speech.

EMG Results

The EMG patterns during speech gestures are distinct from the patterns associated with non-speech maneuvers. Deep and superficial masseter, and anterior and posterior temporalis were never active in a manner consistently related to jaw movement during speech, even when non-speech activity in these muscles was substantial. The four remaining muscles generally showed consistent activity during speech gestures, medial pterygoid and superior lateral pterygoid associated with raising the jaw, inferior lateral pterygoid and anterior belly of digastric lowering the jaw.

Table 1 presents muscle activity for speech gestures expressed as a percentage of the same muscle's non-speech maximum. It should be noted that the individual differences are very large. To assess whether these individual differences are a function of varying electrode placements, we repeated the experiment with Subject VR. From Table 1 it is apparent that the ratio of muscle activity for speech and non-speech gestures differed from those observed during the first experimental session. However, the pattern of activity among speech gestures was basically consistent from session to session. The pattern of activity across sessions changed only for the inferior head of lateral pterygoid. In the first recording session, activity in this muscle was distributed among the vowel articulations in a pattern different from that occurring for all other subjects. In the second session, the pattern of activity in the inferior head of lateral pterygoid was similar to the pattern for all other subjects. Thus, in the data analyses presented below, the absolute values of muscle activity are not crucial and are likely to change across recording sessions. What appears to be consistent is the pattern of muscle activity distributed among speech gestures.

The level of muscle activity for speech gestures was generally highest either when raising the jaw from the open vowel /a/ to the following consonant constriction or when lowering the jaw from the consonant constriction to the open position for /a/. In order to clarify presentation of the results, the levels of activity reported below represent these maximal movements. For levels of activity corresponding to jaw movements to and from /i/ and /u/ refer to Table 1.

Medial pterygoid activity associated with speech gestures was observed for three speakers. Data for one of these speakers are presented in Figure 2. Medial pterygoid activity associated with elevating the jaw from its position for /a/ reached 21% (Subject CC) and 40% (Subject BT) of the maximum activity

Figure 1. (a) Averaged EMG activity of the inferior head of the lateral pterygoid muscle and (b) vertical mandibular movement for subject CC.

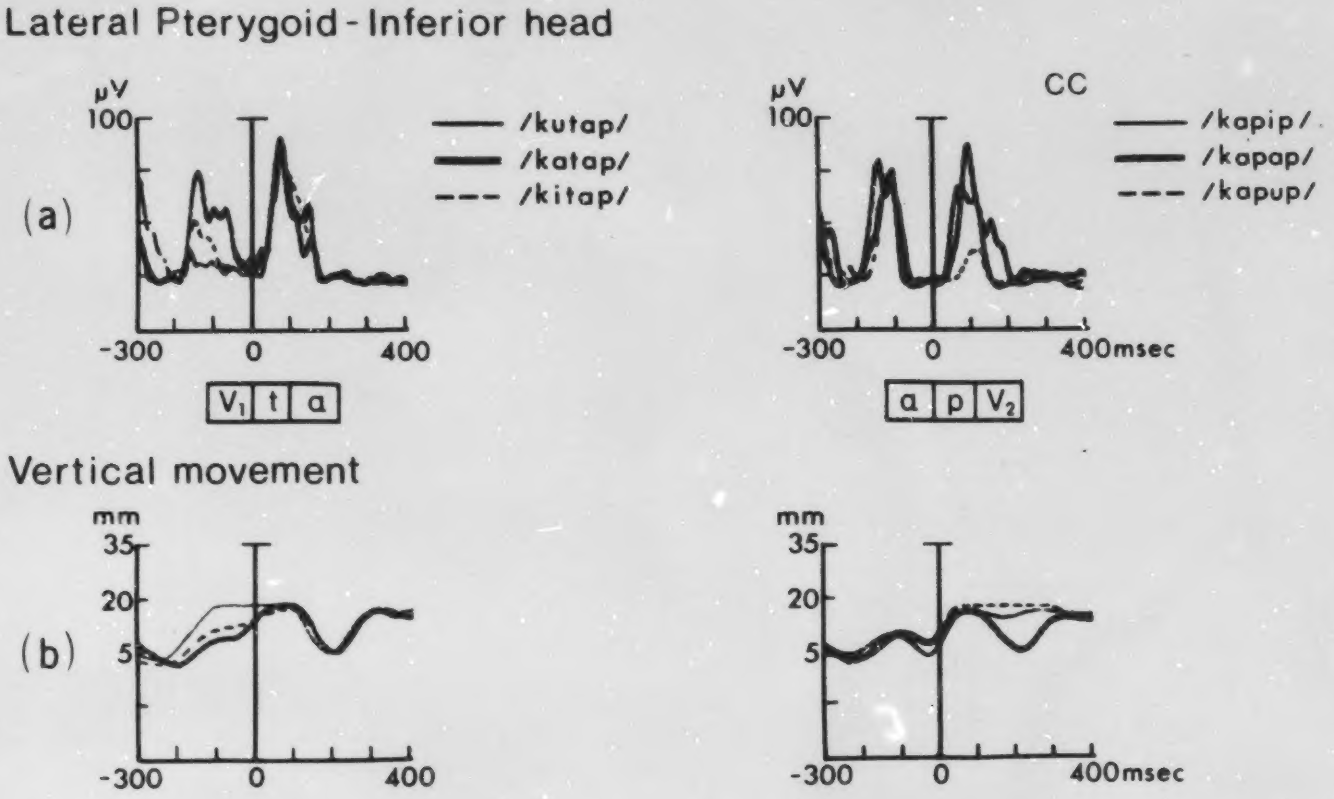


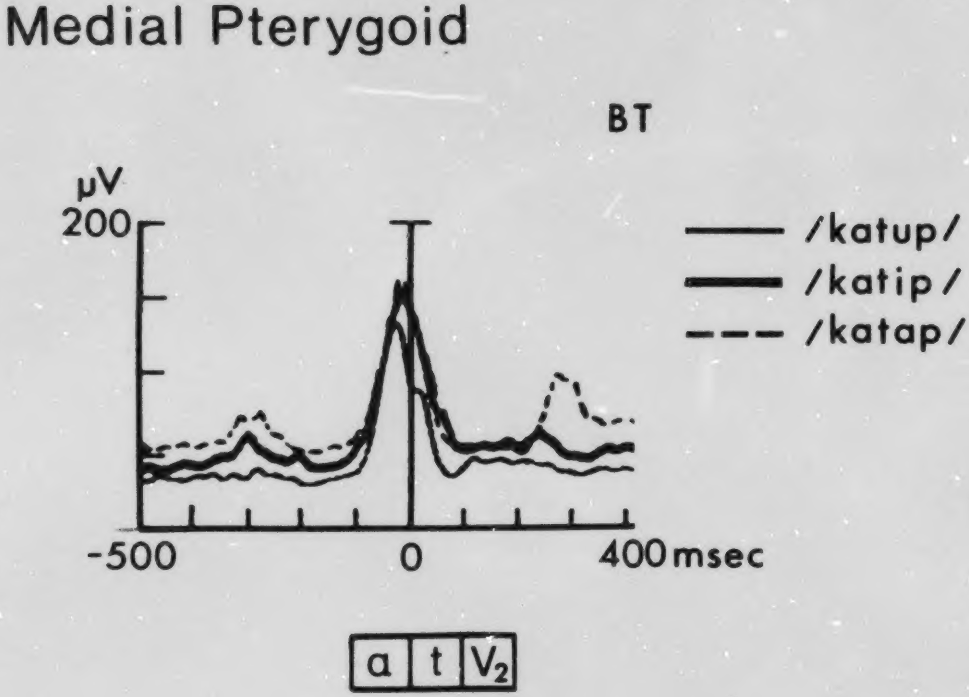
Table 1

Muscle activity for speech gestures expressed as a percentage of each muscle's non-speech maximum. For the jaw-raising muscles, values represent activity recorded when the jaw moved from the vowel indicated to the following consonant. For the jaw-lowering muscles, values represent activity recorded when moving from the consonant constriction to the following vowel.

		Muscle			
		Medial Pterygoid	Lateral Pterygoid Superior	Lateral Pterygoid Inferior	ABD
S	Vowel				
BE	/a/	-	178	26	53
	/i/	-	154	15	30
	/u/	-	130	15	30
BT	/a/	40	35	100	-
	/i/	21	10	100	-
	/u/	17	9	41	-
CC	/a/	21	*	89	36
	/i/	21	*	79	37
	/u/	17	*	72	9
VR$_1$	/a/	130	61	0	184
	/i/	84	54	69	141
	/u/	78	56	102	126
VR$_2$	/a/	45	69	31	63
	/i/	41	61	22	34
	/u/	37	58	21	32

-activity not specific to the gesture
* bad electrode insertion

Figure 2. Averaged EMG activity of the medial pterygoid muscle for subject BT.



for non-speech elevation of the jaw and clenching the teeth. In contrast, for Subject VR peak medial pterygoid activity associated with moving from /a/ to the following consonant constriction reached 130% of its non-speech maximum activity. Thus, there was no consistent relationship between the maximum activity for speech and non-speech gestures.

The superior head of lateral pterygoid was active during elevation of the mandible for all speakers. Figure 3 presents these data for one subject. When viewed in relation to maximum activity in the superior head of lateral pterygoid during the non-speech gestures of elevating and clenching the mandible, peak activity for moving from /a/ to the following consonant constriction reached 178%, 35%, and 61% for speakers BE, BT, and VR, respectively. The relationship between activity in superior lateral pterygoid during speech and non-speech shows no consistent pattern across speakers.

The inferior head of lateral pterygoid was active for the production of the vowels, presumably to lower the mandible (Figure 3). This activity was also examined in relation to the maximum activity achieved during non-speech gestures. For Subject BE, peak activity in the inferior head of lateral pterygoid when raising the jaw from /a/ was only 26% of the non-speech maximum level. For Subject BT, peak inferior lateral pterygoid activity associated with speech and non-speech gestures was identical: Peak activity for /a/ was 100% of the maximum achieved for mandibular protrusion. For Subject CC, peak activity in the inferior head of lateral pterygoid reached 89% of the non-speech maximum level. Again, the relationship of muscle activity in speech and non-speech gestures is inconsistent across speakers.

The pattern of activity in inferior lateral pterygoid is distributed differently among the vowel articulations of Subject VR in that the level of activity is near base line for production of /a/ but high for production of /i/ and /u/. The only case in which the inferior head of lateral pterygoid was active for production of a consonant was for the production of /f/ in /af/ by Subject VR, possibly reflecting a protrusive component of the elevating gesture.

Activity in the anterior belly of the digastric (ABD) was seen during speech in only three of the four speakers. The maximum level of activity in ABD was associated with jaw lowering for production of the open vowel /a/. One speaker's data are presented in Figure 4. The magnitude of the muscle activity was assessed as a percent of the maximum activity recorded in ABD during non-speech gestures, in this case large-excursion lowering of the mandible. For Subject BE, peak ABD activity achieved for production of /a/ reached 53% of the maximum activity of ABD for non-speech jaw lowering. For Subject CC, maximum ABD activity during speech was 36% of the maximum activity during non-speech gestures. In contrast, for Subject VR peak ABD activity associated with the production of /a/ was 184% of the maximum activity in ABD during non-speech lowering of the jaw.

The basic pattern that emerges is that medial pterygoid and superior lateral pterygoid act in relation to raising the jaw, whereas inferior lateral pterygoid and anterior belly of the digastric function to lower the jaw. Note that for each subject the two heads of lateral pterygoid functioned as two separate muscles: The inferior head was active during jaw lowering, the

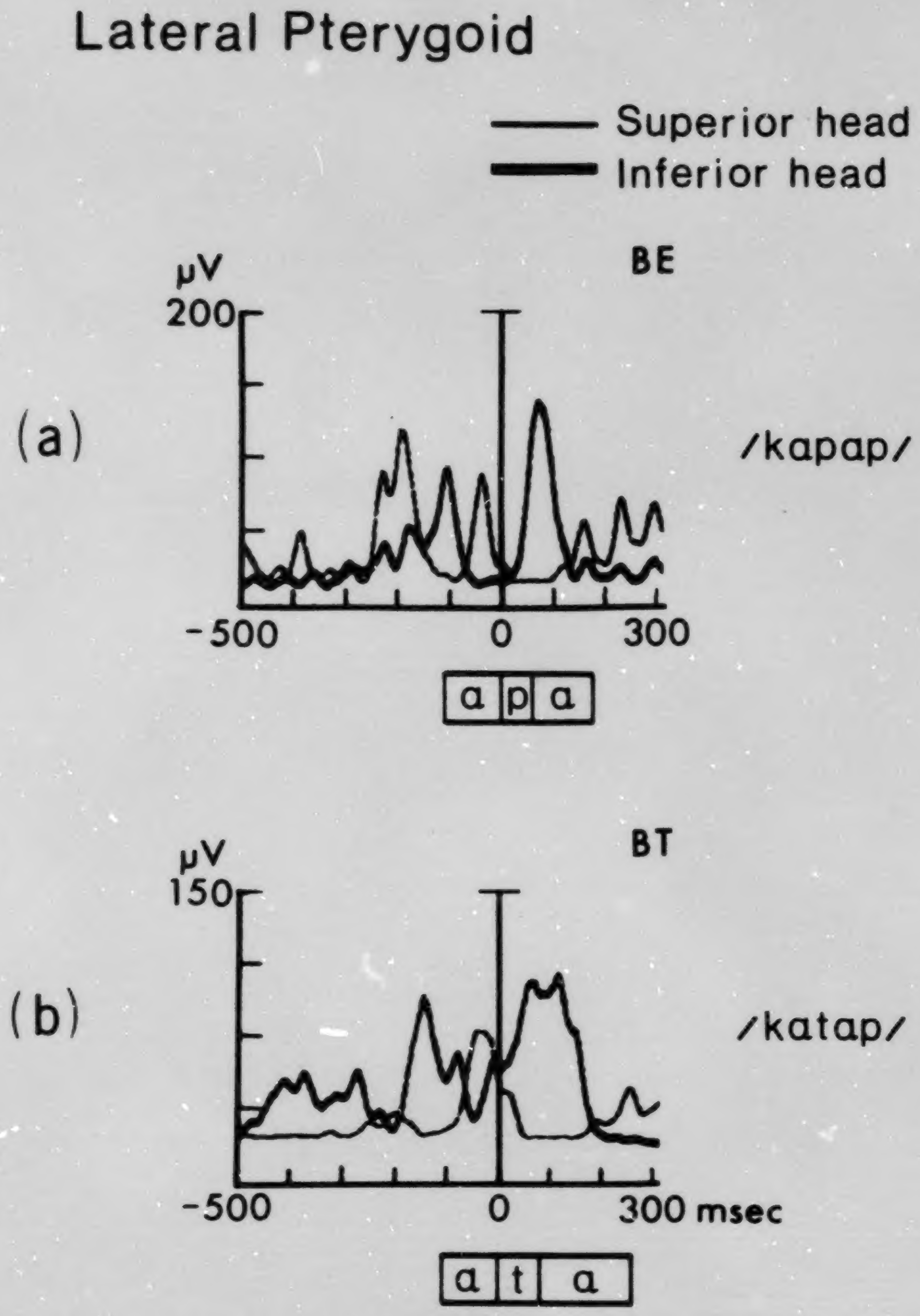


Figure 3. Averaged EMG activity of the superior and inferior heads of the lateral pterygoid muscle for (a) subject BE and (b) subject BT.

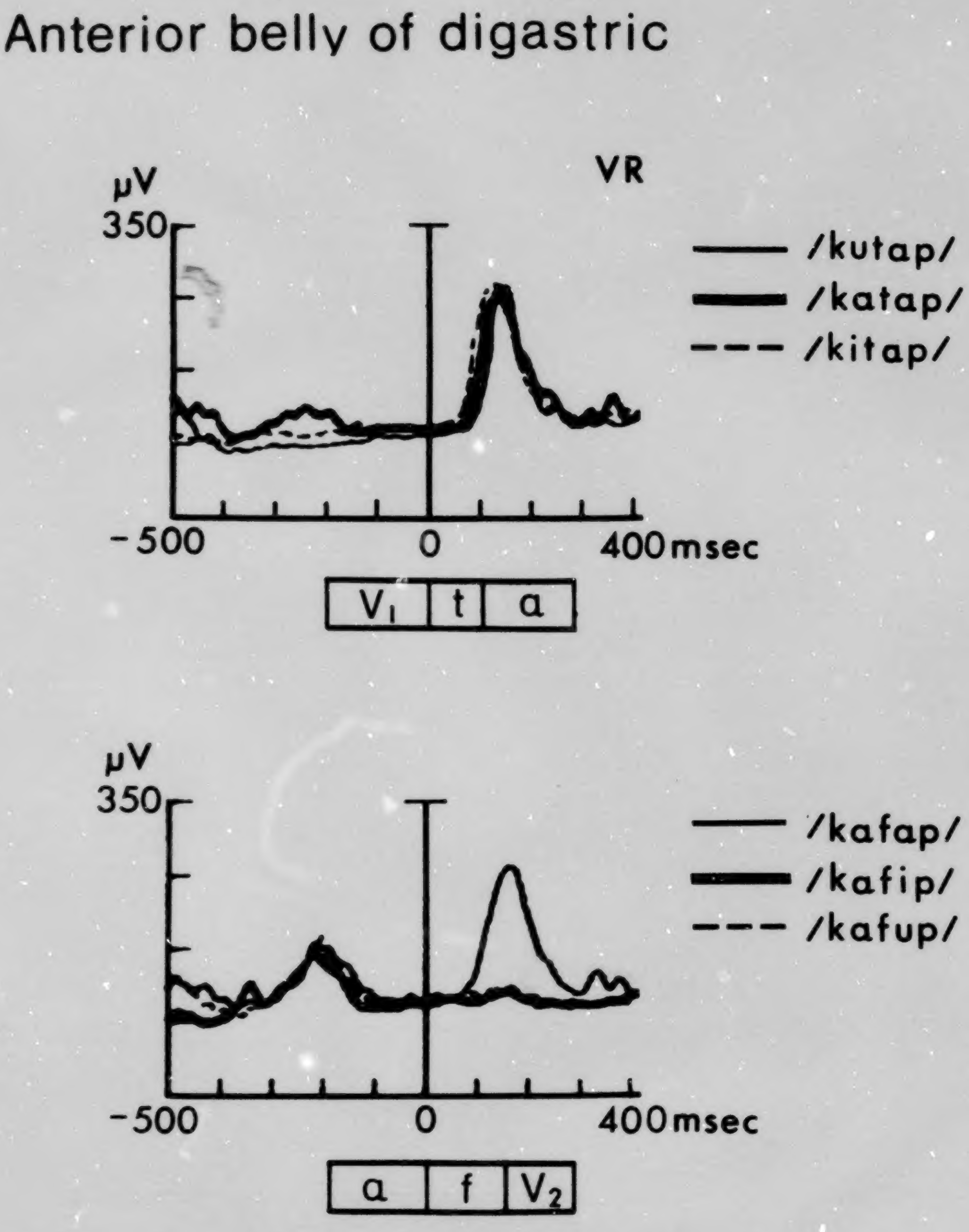


Figure 4. Averaged EMG activity for the anterior belly of the digastric muscle for subject VR.

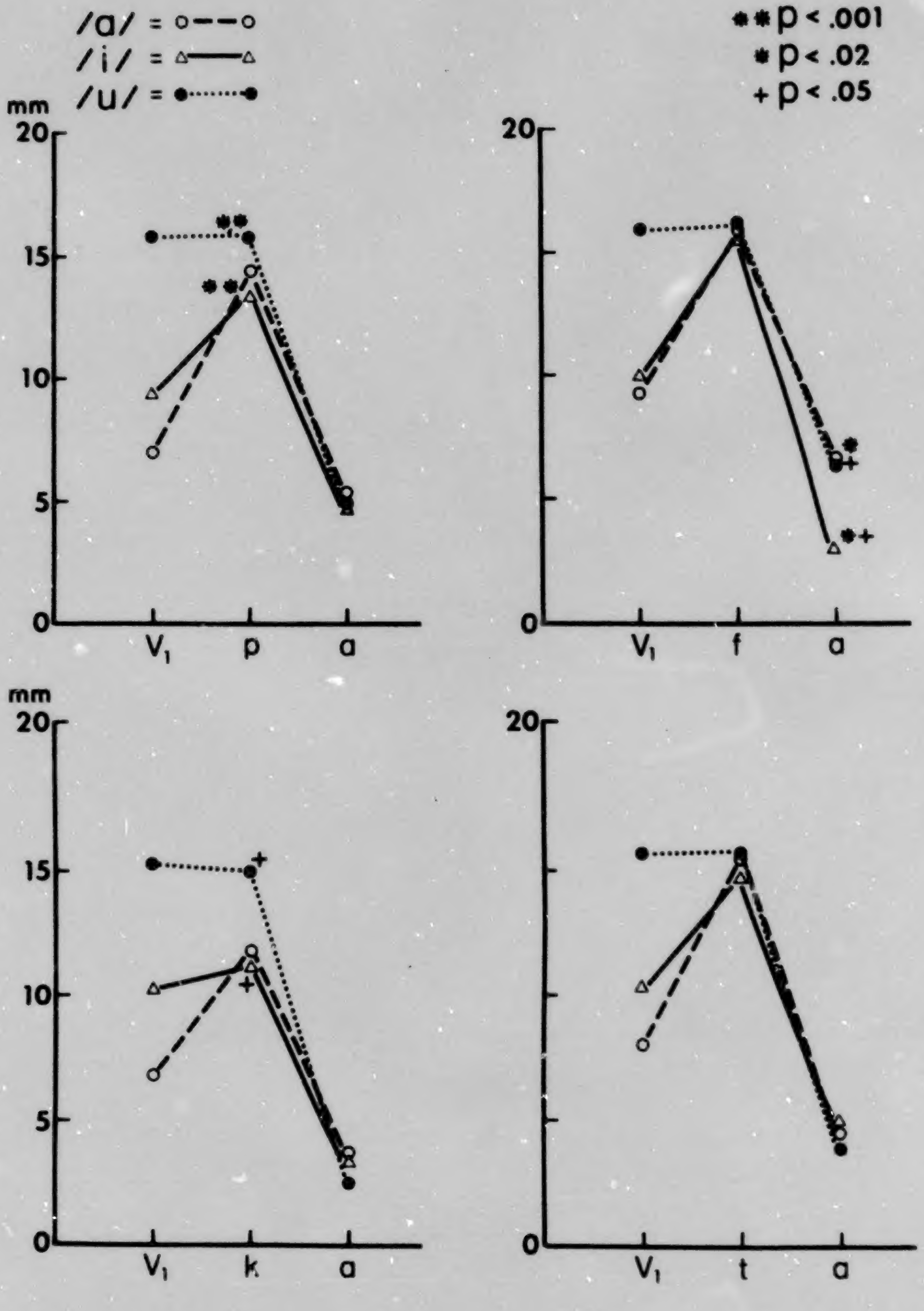


Figure 5. Vertical movement of the mandible is represented (in mm) for subject CC. Movement curves were averaged over V_1, holding C and V_2 constant.

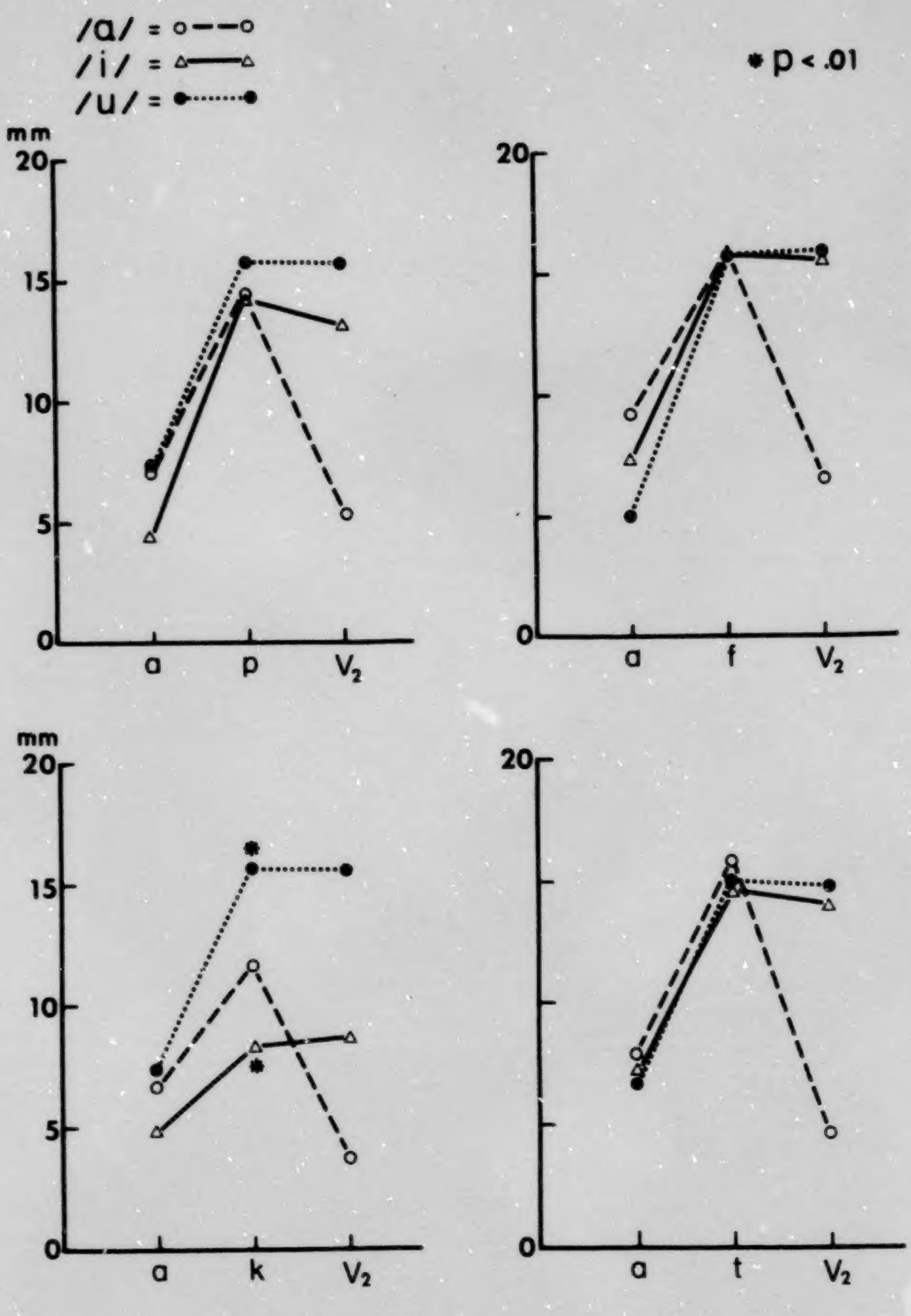


Figure 6. Vertical movement of the mandible for subject CC. Curves represent averages pooled over V_2, holding C and V_1 constant.

superior head during jaw elevation. The superior head of lateral pterygoid has been thought to stabilize the mandibular condyles, particularly when the upper and lower teeth are in contact. When speakers produce the speech sample used in this study it is likely that the teeth are rarely, if ever, in contact. However, consistent superior lateral pterygoid activity occurred in relation to the gesture of jaw elevation. Thus, superior lateral pterygoid has an elevating function during speech rather than the stabilizing function normally ascribed to it. The activity patterns of the inferior and superior heads were basically reciprocal and within each subject were patterned differently for each vowel gesture.

Coarticulation--Results and Discussion

The data reported here may also be used to examine coarticulatory effects on mandibular displacement and its underlying muscle activity. Mandibular displacement for the consonant was examined in relation to both the preceding and following vowels (Figures 5 and 6). Production of /f/ and /t/ was insensitive to changes in V_1 or V_2. The position of the mandible for /p/ showed no anticipation of V_2 but varied with changes in V_1. Mandibular displacement for /k/ showed both carryover and anticipatory effects of the preceding ($t=3.81$, $p<.05$) and following ($t=7.70$, $p<.01$) vowels.

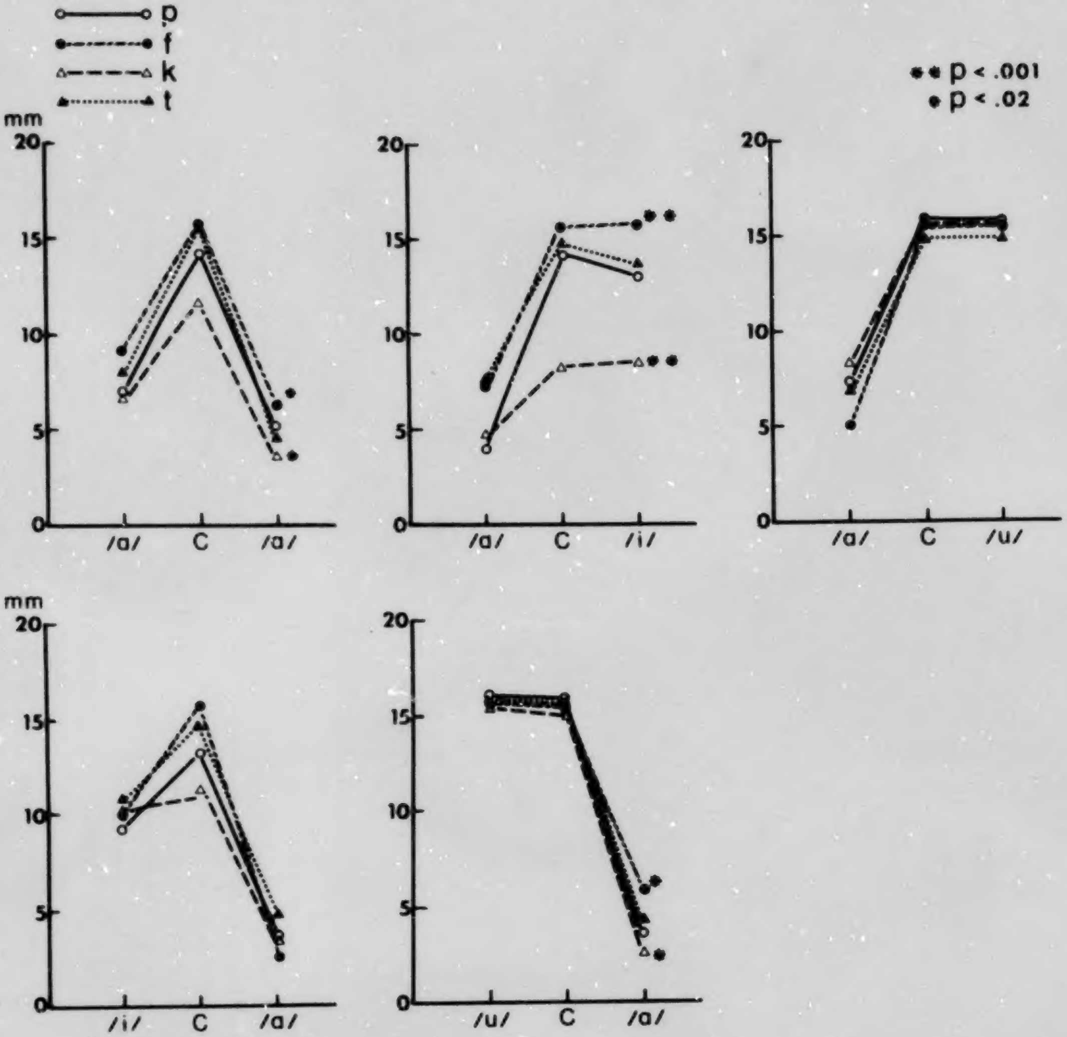
Variations in mandibular height during vowel production were examined as a function of the intervocalic consonant. In no case was mandibular position for V_1 effected by changes in the following consonant; that is, there was no evidence of anticipatory changes in mandibular height (Figure 7). Jaw position for V_2 was examined as a function of the preceding consonant. In symmetrical utterances /aCa/ jaw position for V_2 was significantly affected by the consonant ($t=5.64$, $p<.01$), resulting in mandibular height for the vowel of the order /ka/, /ta/, /pa/, /fa/ (lowest to highest; Figure 7). These carryover effects of the consonant on V_2 were also evident in mandibular displacement for /aCi/ (/ki/ > /pi/ > /ti/ > /fi/, $t=10.14$, $p<.001$), and /uCa/ (/ka/ > /pa/ > /ta/ > /fa/, $t=4.86$, $p<.01$; Figure 7). Utterances of the form /aCu/ and /iCa/ showed no carryover effect of C on V_2. Again, in no case were anticipatory effects evident.

Jaw position during vowel production was examined as to whether it was effected by the other vowel in the sequence (Figure 1). In all cases, jaw position for V_1 never anticipated V_2. Jaw position for V_2 showed carryover effects of V_1 only when the intervocalic consonant was /f/. However, the changes in jaw position were complex and, unlike the results reported by Gay (1974), do not simply reflect or invert jaw height for V_1. For all other intervocalic consonants, jaw position for V_2 was insensitive to jaw position for V_1.

The EMG activity was examined for vowel-to-vowel coarticulatory effects. For those muscles that showed consistent EMG activity in speech there was no carryover effect of V_1 on V_2 or anticipatory effect of V_2 on V_1 (see Figures 1, 2, and 4).

In sum, although the trajectory of the mandible during the speech utterances used in this study are in basic agreement with those noted by Perkell (1969) and Gay (1974), less coarticulatory influence was evident than

Figure 7. Movement of the mandible in the vertical dimension for subject CC. Curves are pooled over consonants, holding V_1 and V_2 constant.



found by Gay (1974). We extended Ohman's (1966) hypothesis of vowel-to-vowel coarticulation by suggesting that EMG activity and the resulting jaw displacement for any given vowel might be dependent on the other vowel in a VCV sequence. However, neither EMG activity related to jaw lowering for V_1 nor jaw position achieved for V_1 was observed to anticipate V_2. EMG activity related to V_2 was not affected by V_1; actual jaw position for V_2 was sensitive to V_1 only when the intervocalic consonant was /f/.

GENERAL DISCUSSION

The data reported here support the traditional classification of masseter, temporalis and medial pterygoid as jaw elevators, and the anterior belly of digastric and inferior lateral pterygoid as jaw depressors, for the performance of non-speech maneuvers. However, this classification does not apply when the activity concerned is speech. No consistent relationship was found between mandibular movement for the speech sounds used in this study and activity of anterior or posterior temporalis, or superficial or deep masseter. However, superficial masseter may be associated with production of phonetic segments for which the jaw is assumed to be in a more protruded position than those examined here (e.g., s, sh). (Ewan, personal communication; Sussman et al., 1973; Tuller, Harris, & Gross, unpublished data).

For the limited inventory of speech gestures reported here, medial pterygoid and superior lateral pterygoid are active during jaw elevation while inferior lateral pterygoid and anterior belly of the digastric act during jaw depression. The functional differentiation of inferior and superior lateral pterygoid for both speech and non-speech gestures agrees well with the data obtained by Grant (1973) and McNamara (1973). The inferior lateral pterygoid functions during lowering of the jaw whereas the superior lateral pterygoid functions during raising of the jaw.

Notable, however, are the many individual differences among speakers, differences not consistent across speech and non-speech gestures. Although this variability may reflect structural differences among speakers, Angle's method of classifying mandibular occlusion does not allow one to predict the pattern of EMG activity in the jaw muscles. Differences in the structure of the temporomandibular joint and/or the angle and placement of muscle attachments may also increase the variability in muscle activity across speakers. Given the complex nature of the temporomandibular joint, the "jaw lowering gesture" in different speakers may consist of different relative amounts of condylar translation and rotation, resulting in different EMG patterns across speakers. Moreover, the relationship between speech and non-speech is not fixed; in these data, the vertical displacement for a maximally lowered jaw did not bear a fixed relation to the lowest mandibular position that occurred when the person was speaking.

In conclusion, investigators have recognized the need for simultaneous recording of different types of information when studying labial-mandibular or lingual-mandibular coordination, during normal or disrupted speech. One source of information is the muscle activity underlying mandibular movement. This study suggests that when studying muscle activity related to jaw raising during speech, the investigator should monitor medial pterygoid and superior lateral pterygoid. The study of electromyographic activity during jaw lowering in speech gestures should include inferior lateral pterygoid and anterior belly of digastric. The pattern of muscle activity during speech may illuminate the nature of speech motor control, but only if the monitored muscles are indeed those directly relevant to the movement.

REFERENCE NOTES

1. Folkins, J. W., Zimmerman, G. N., & Cooper, D. S. *Activity in jaw closing muscles during speech*. Paper presented to ASHA, San Francisco, California, November, 1978.
2. Gross, B. D., & Lipke D. *A technique for percutaneous lateral pterygoid electromyography*. Manuscript submitted for publication.

REFERENCES

Ahlgren, J. Mechanism of mastication. A quantitative cinematographic and electromyographic study of masticatory movements in children, with special reference to occlusion of the teeth. *Acta Odontologica Scandinavica*, 1966, *24*, Suppl. 44.

Carlsoo, S. Nervous coordination and mechanical function of the mandibular elevators. *Acta Odontologica Scandinavica*, 1952, *10*, Suppl. 11.

Carlsoo, S. An electromyographic study of the activity and an anatomical analysis of the mechanism of the lateral pterygoid muscle. *Acta Anatomica* (Basel), 1956, *26*, 339-351.

Dubner, R., Sessle, B. J., & Storey, A. T. *The neural basis of oral and facial function*. New York: Plenum Press, 1978.

Ewan, W. G., & Krones, R. Measuring larynx movement using the thyroumbrometer. *Journal of Phonetics*, 1974, *2*, 327-335.

Folkins, J. W., & Abbs, J. H. Lip and jaw motor control during speech: Responses to resistive loading of the jaw. *Journal of Speech and Hearing Research*, 1975, *18*, 207-221.

Gay, T. Jaw movements during speech: A cinefluorographic investigation. *Haskins Laboratories Status Report on Speech Research*, 1974, *SR-39/40*, 219-229.

Gay, T., & Turvey, M. T. Effects of efferent and afferent interference on speech production: Implications for a generative theory of speech motor control. In *Proceedings of the Ninth International Congress of Phonetic Sciences*, 1979, *Vol. II*, 344-350.

Gibbs, C. H., & Messerman, T. Jaw motion during speech. In *Orofacial function: Clinical research in dentistry and speech pathology*. (ASHA Reports, No. 7). Washington, D.C.: American Speech and Hearing Association, 1972.

Grant, P. Lateral pterygoid: Two muscles? *American Journal of Anatomy*, 1973, *138*, 1-9.

Griffin, C. J., & Malor, R. An analysis of mandibular movement. In Y. Kawamura (Ed.), *Physiology of mastication*. Basel: S. Karger, 1974.

Griffin, C., & Munro, R. EMG of the jaw-closing muscles in the open-close-clench cycle in man. *Archives of Oral Biology*, 1969, *14*, 141-149.

Hamlet, S. L., & Stone, M. Compensatory vowel characteristics resulting from the presence of different types of experimental dental prostheses. *Journal of Phonetics*, 1976, *4*, 199-218.

Hamlet, S. L., & Stone, M. Compensatory alveolar consonant production induced by wearing a dental prosthesis. *Journal of Phonetics*, 1978, *6*, 227-248.

Hickey, J. C., Stacy, R. W., & Rinear, L. L. Electromyographic studies of mandibular muscles in basic jaw movements. *Journal of Prosthetic Dentistry*, 1957, *7*, 565-570.

Hirose, H. Electromyography of the articulatory muscles: Current instrumentation and techniques. *Haskins Laboratories Status Report on Speech Research*, 1971, *SR-25/26*, 73-86.

Kerr, D. A., Ash, M. M., & Millard, H. D. *Oral diagnosis* (5th ed.). St. Louis: C. V. Mosby and Co., 1978.

Kewley-Port, D. Computer processing of EMG signals at Haskins Laboratories. *Haskins Laboratories Status Report on Speech Research*, 1973, *SR-33*, 173-184.

Kewley-Port, D. An experimental evaluation of the EMG data processing system: Time constant choice for digital integration. *Haskins Laboratories Status Report on Speech Research*, 1974, *SR-37/38*, 65-72.

Lindblom, B., Lubker, J., & Gay, T. Formant frequencies of some fixed-mandible vowels and a model of speech motor programming. *Journal of Phonetics*, 1979, *7*, 147-162.

MacNeilage, P. F., Sussman, H. M., Westbury, J. R., & Powers, R. K. Mechanical properties of single motor units in speech musculature. *Journal of the Acoustical Society of America*, 1979, *65*, 1047-1052.

McNamara, J. A. The independent functions of the two heads of the lateral pterygoid muscle. *American Journal of Anatomy*, 1973, *138*, 197-206.

Møller, E. The chewing apparatus. *Acta Physiologica Scandinavica*, 1966, *69*, (Suppl. 280), 1-229.

Møller, E. Action of the muscles of mastication. In Y. Kawamura (Ed.), *Physiology of mastication*. Basel: S. Karger, 1974.

Moyers, R. E. An electromyographic analysis of certain muscles involved in temporomandibular movement. *American Journal of Orthodontics*, 1950, *37*, 481-515.

Ohman, S. E. G. Coarticulation in VCV utterances: Spectrographic measurements. *Journal of the Acoustical Society of America*, 1966, *39*, 151-168.

Perkell, J. S. (1969) *Physiology of speech production: Results and implications of a quantitative cineradiographic study*. Cambridge, Mass.: MIT Press, 1969.

Sarnat, B. G. *The temporomandibular joint* (2nd ed.). Springfield, Ill.: Thomas, 1964.

Sicher, H., & DuBrul, E. L. *Oral anatomy* (6th ed.). St. Louis: Mosby, 1976.

Sussman, H. M., MacNeilage, P. F., & Hanson, R. J. Labial and mandibular dynamics during the production of bilabial consonants: Preliminary observations. *Journal of Speech and Hearing Research*, 1973, *16*, 397-420.

Woelfel, J. B., Hickey, J. C., Stacy, R. W., & Rinear, L. Electromyographic analysis of jaw movements. *Journal of Prosthetic Dentistry*, 1960, *10*, 688-697.

Yemm, R. The orderly recruitment of motor units of the masseter and temporalis muscles during voluntary isometric contraction in man. *Journal of Physiology (London)*, 1977, *265*, 163-174.

LARYNGEAL ACTIVITY IN SWEDISH VOICELESS OBSTRUENT CLUSTERS[*]

Anders Löfqvist[+] and Hirohide Yoshioka[++]

Abstract. Laryngeal articulatory movements and their coordination with supralaryngeal events have proved to be important for control of voicing and pre- and post-aspiration in obstruents. A reciprocal pattern of activity has generally been observed among laryngeal abductor and adductor muscles in the control of glottal opening area in voiceless obstruent production. Current notions about laryngeal articulatory control rest, however, mainly on studies using simple linguistic materials, where voiced and unvoiced segments alternate in a regular manner. The present study examines laryngeal activity in voiceless obstruent clusters using the combined techniques of electromyography, fiberoptic filming and transillumination of the larynx. The results indicate that laryngeal articulatory movements are organized in one or more continuous opening and closing gestures, which are precisely coordinated with oral articulations to meet the aerodynamic requirements of speech production. Comparison of temporal patterns of glottal area variations obtained by fiberoptic filming and by transillumination of the larynx, showed them to be practically identical, which was taken as positive evidence for the use of transillumination in speech research.

INTRODUCTION

Technical developments in recent years have provided means for a better understanding of laryngeal activity in speech. Application of electromyographic, fiberoptic and glottographic techniques has advanced our knowledge of the control of laryngeal articulatory movements and their coordination with supralaryngeal events in speech production. In particular, the important role of the larynx and of laryngeal-oral coordination for producing contrasts of voicing and of pre- and post-aspiration in obstruents have been clarified for several different languages (Löfqvist, in press).

Glottal opening during voiceless obstruent production has been shown to be controlled by the posterior cricoarytenoid (abduction) and the interarytenoid (adduction) muscles (Hirose, 1976; Hirose, Yoshioka, & Niimi, 1978). For these two muscles, a reciprocal pattern of activation has generally been

---

[*] Parts of this work were presented at the 97th Meeting of the Acoustical Society of America, 11-15 June 1979, Cambridge, Mass.

[+] Also Lund University, Lund, Sweden.

[++] Also University of Tokyo, Tokyo, Japan.

Acknowledgment: This work was supported by NINCDS Grants NS-13617 and NS-13870, and by NIH Grant RR-5596.

[HASKINS LABORATORIES: Status Report on Speech Research SR-59/60 (1979)]

observed, whereas the role of other adductor muscles, such as the lateral cricoarytenoid, is considerably more unclear. The lateral cricoarytenoid is usually suppressed for both voiced and voiceless obstruents, and has been functionally grouped with the vocalis muscle (Hirose & Gay, 1972; Hirose & Ushijima, 1978). It should be noted, however, that the notion of strict reciprocity between the posterior cricoarytenoid and the interarytenoid muscles rests mainly on studies using combinations of consonants and vowels, where voiced and unvoiced segments alternate in a regular manner.

Laryngeal activity in clusters of voiceless obstruents has not been dealt with in any comparable detail. Fujimura and Sawashima (1971) investigated a limited number of voiced and voiceless stop combinations in American English using fiberoptic filming. Löfqvist (1977, 1978) studied Swedish voiceless obstruent clusters using transillumination of the larynx and aerodynamic records. Pétursson (1978) applied the same techniques to Icelandic obstruent combinations.

The results of the Swedish and Icelandic studies indicated that a sequence of two voiceless obstruents could be produced with one or two separate glottal opening and closing gestures. Combinations of voiceless stop + voiceless fricative, or voiceless fricative + voiceless unaspirated stop generally contained only one glottal articulatory gesture, with peak glottal opening occurring during the fricative. Similarly, two consecutive voiceless stop consonants were produced with one laryngeal gesture, the timing of which varied with the presence or absence of aspiration after the release of the second stop. On the other hand, a sequence of voiceless fricative + voiceless aspirated stop usually contained two separate laryngeal gestures with peak glottal opening during the fricative and just before stop release; between the two maxima of glottal opening the vocal folds were adducted without complete glottal closure.

Due to the limited number, and scope, of investigations using more complex linguistic material, it seems important to determine if established notions about laryngeal function also are valid for laryngeal control in clusters of voiceless obstruents, where the control of the larynx, in terms of laryngeal "coarticulation," may differ from that in single obstruents. In doing so it is, moreover, important to obtain simultaneous information on both glottal articulatory movements and the activity of the muscles assumed to be responsible for these movements. In the absence of such information, it may be difficult to determine either the specific effects of different muscular activity patterns, or whether observed movements are actually caused by muscular and/or nonmuscular, e.g., aerodynamic, forces.

The transillumination technique, or photoglottography (Sonesson, 1960), has been in regular use for several years, but it has not been extensively compared with other techniques for obtaining information on laryngeal behavior in speech. Such comparisons have mostly involved transillumination and electrical glottography during phonation (e.g., Frøkjaer-Jensen, 1968; Köster & Smith, 1970; Kitzing, 1977; Kitzing & Löfqvist, 1978). Two studies of phonation, comparing variations in glottal area measured from high-speed films and transillumination signals (Coleman & Wendahl, 1968; Harden, 1975), reported conflicting evidence concerning the reliability of the transillumination technique. No study has compared information on dynamic patterns of laryngeal articulations obtained by transillumination and by other methods. Since this is an area where the transillumination technique may be most useful, it appears important to evaluate its possibilities and limitations for this kind of research.

The aim of the present study is twofold: To contribute to a better understanding of the biomechanics of laryngeal control in speech, and to further assess the transillumination technique as a tool for studying laryngeal behavior in speech.

We have thus studied laryngeal activity in Swedish obstruent clusters using the combined techniques of electromyography, fiberoptic filming and transillumination of the larynx. The results show that laryngeal articulatory movements are organized in one or more continuous opening and closing gestures, which are precisely coordinated with supralaryngeal events to meet the aerodynamic requirements for producing a signal with a specified acoustic structure. The results also indicate that the notion of strictly reciprocal activity between the posterior cricoarytenoid and the interarytenoid muscles may need some qualification and further refinement. As to the second objective, temporal patterns of variations in glottal opening area obtained by fiberoptic filming and by transillumination proved to be practically identical; we may thus conclude that transillumination is a viable technique in studies of laryngeal activity in speech.

METHOD

Procedure

Electromyographic recordings were obtained from the posterior cricoarytenoid (PCA) and the interarytenoid (INT) muscles. Bipolar, hooked-wire electrodes (Basmajian & Stecko, 1962; Hirano & Ohala, 1969), consisting of a pair of platinum-tungsten alloy wires (50 microns in diameter with isonel coating), were inserted perorally under indirect laryngoscopy with the aid of a specially designed, curved probe. Before the insertion, surface anesthesia (4% Xylocain) was applied to the pharyngeal and laryngeal mucosa. The EMG signals were recorded and processed with the Haskins Laboratory system (Kewley-Port, 1977). After amplification and high-pass filtering at 80 Hz, to remove movement artifacts and hum, the signals were recorded on a multichannel instrumentation tape recorder (Consolidated Electrodynamics, VR-3300). For processing, the signals were full wave rectified and integrated over a 5 msec window through linear-reset integrators, and fed into a DDP-224 computer at a sampling rate of 200 Hz. In the averaging process, the signals were aligned with reference to a predetermined, acoustically defined line-up point, and also further integrated over 35 milliseconds.

The larynx was filmed through a flexible fiberscope (Olympus VF Type 0) at a film speed of 60 frames/second. The fiberscope, inserted through the nose, was kept in position by a specially designed headband. A synchronization signal was recorded on one channel of the tape recorder for frame identification. Relevant portions of the film were analyzed frame by frame with a computer assisted analyzing system, and the distance between the vocal processes was measured as an index of glottal opening.

The light from the fiberscope was used as part of a transillumination system, whereby the amount of light passing through the glottis was sensed by a phototransistor (Philips, BPX 81), placed on the surface of the neck just below the cricoid cartilage, and held in position by a neckband. The signal from the transistor was amplified and recorded on one channel of the tape recorder. For averaging, the transillumination signal was sampled at 200 Hz and fed into the computer. It was aligned with the EMG signals and integrated over 5 milliseconds.

The measurements from the film were combined with the transillumination signals obtained for the same tokens of the test utterances. The line-up points for the film material were decided visually and marked by hand. During this process, some misalignments may possibly have occurred, since the temporal resolution between adjacent frames was 17 milliseconds. No further processing was applied to the measurements from the film.

A direction-sensitive microphone was used to record the audio signal in direct mode on one channel of the instrumentation recorder and also on an ordinary tape recorder at a recording speed of 7.5 ips. The audio signal was used for determination of the line-up points and also included in the further processing. It was sampled at 10 kHz using the Haskins PCM system and then rectified and analyzed in parallel with the bioelectrical and biomechanical signals. In the averaging process the rectified audio signal was integrated over 15 milliseconds.

Linguistic material

The linguistic material consisted of Swedish voiceless obstruents and obstruent clusters in various positions, with a word boundary preceding, following or intervening within the cluster. Both the transillumination technique and fiberoptic filming require a wide pharyngeal cavity, which had to be taken into account in selecting the linguistic material. Swedish words were used, and these words are given in Table 1. All the words in Set A were combined with those in Set B and placed in the carrier phrase "Men ..." ("But ...") to yield 24 normal Swedish sentences.

Table 1

The linguistic material. All the words in set A were combined with those in set B. The words in set A are proper names, those in set B present tense verb forms.

Set A		Set B	
Li	[li:]		
Lis	[li:s]	ilar	[i:lar]
Ek	[e:k]	silar	[si:lar]
Liszt	[lIst]	pilar	[p^{h}i:lar]
Eks	[e:ks]	spelar	[spe:lar]
Kvists	[k^{h}vIsts]		

Swedish voiceless stops are aspirated in prestress position and unaspirated when they immediately follow a stressed vowel or /s/. Although this difference between aspirated and unaspirated voiceless stops is not phonemic in Swedish, when aspiration occurs it serves as one of the cues for the distinction between voiced and voiceless stops, since the former are always unaspirated. In addition, the presence or absence of aspiration in voiceless stops in some contexts marks the location of a word boundary. Word initial stressed vowels in Swedish are usually produced with a glottal attack at the onset.

A native male speaker read the whole material 20 times from randomized lists. Ten to fifteen repetitions of each utterance type were used for averaging. Fiberoptic films were made during 2-5 of the repetitions.

RESULTS

Transillumination and fiberoptics

Patterns of glottal area variations measured from fiberoptic films and by transillumination are shown in Figures 1 and 2. Figure 1 presents the utterances "Liszt ilar," "Liszt silar," "Liszt pilar," and "Liszt spelar," and Figure 2 the utterances "Ek ilar," "Ek silar," "Ek pilar," and "Ek spelar." Two repetitions of each utterance type are shown. The movement patterns obtained by the two techniques were practically identical. This was also shown by a correlation analysis applied to the two curves. For all the utterances observed (n=56), the correlation was highly significant ($p<0.001$).

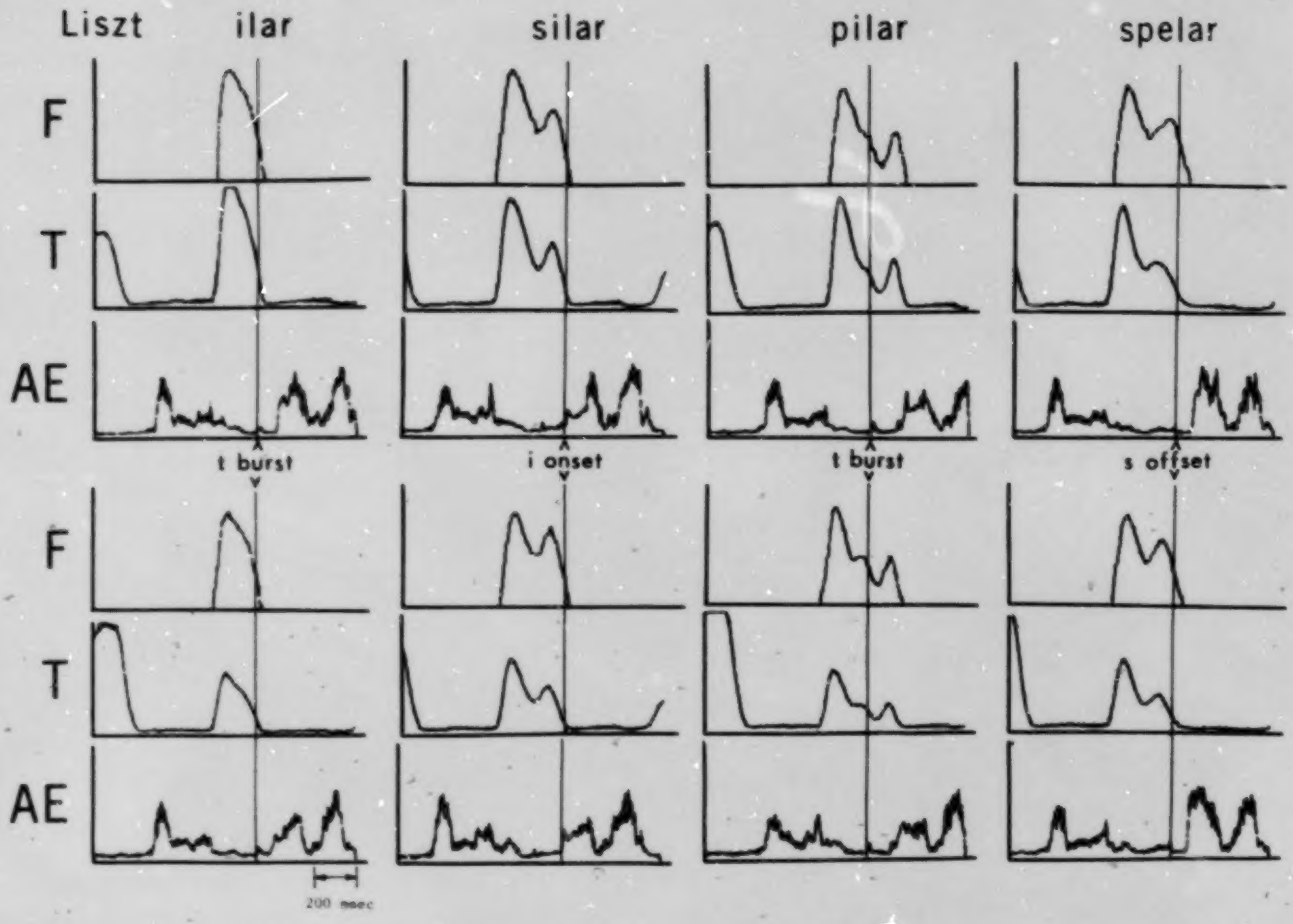
In some instances articulatory movements of the root of the tongue and the epiglottis interfered with the passage of light from the fiberscope to the larynx, cf. "Ek spelar" in Figure 2. These instances could be readily identified by a sudden decrease in the amplitude of the transillumination signal, which lacked any counterpart in the measurements from the film. Inspection of the corresponding film frames indicated that in these cases the view of the anterior portion of the glottis was blocked.

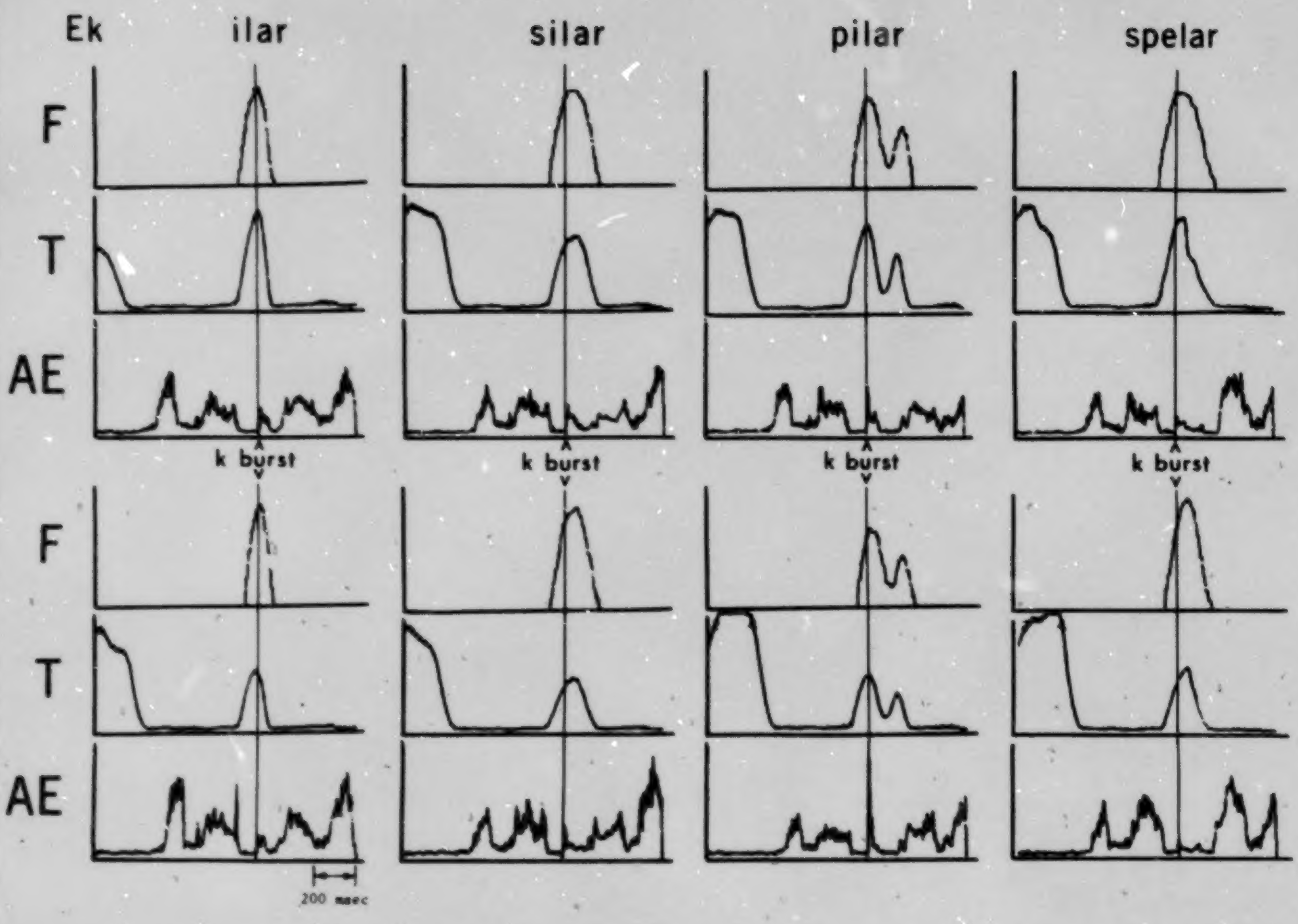
Laryngeal articulatory movements

Since the temporal patterns of glottal area variations obtained by transillumination and by fiberoptic filming showed a high correlation and were practically identical, only those obtained by the former method will be discussed below. The variability among individual tokens of the same utterance type was rather small as shown in Figure 3. The most obvious variation in Figure 3 is that /s#s/ in "Kvists silar" is produced with a single glottal opening gesture, with or without an extra adjustment for maintaining an open glottis throughout the period of frication. We shall therefore focus our attention mainly on the average curves.

In single voiceless obstruents, shown in Figure 4, laryngeal articulatory movements usually have the form of a single "ballistic" opening and closing gesture. Some variation in this gesture can be found for fricatives and aspirated stops. In particular, peak glottal opening occurs closer to implosion for the fricative than for the stop. Glottal abduction also occurs at higher velocity, and peak glottal opening is somewhat larger for the

Figure 1. Comparisons of fiberoptic and transillumination records for 2 tokens of 4 utterance types. F = glottal area obtained by fiberoptic filming. T = glottal area obtained by transillumination. AE = audio envelope.





igure 2. Comparisons of fiberoptic and transillumination records for 2 tokens of 4 utterance types. Symbols as in Figure 1.

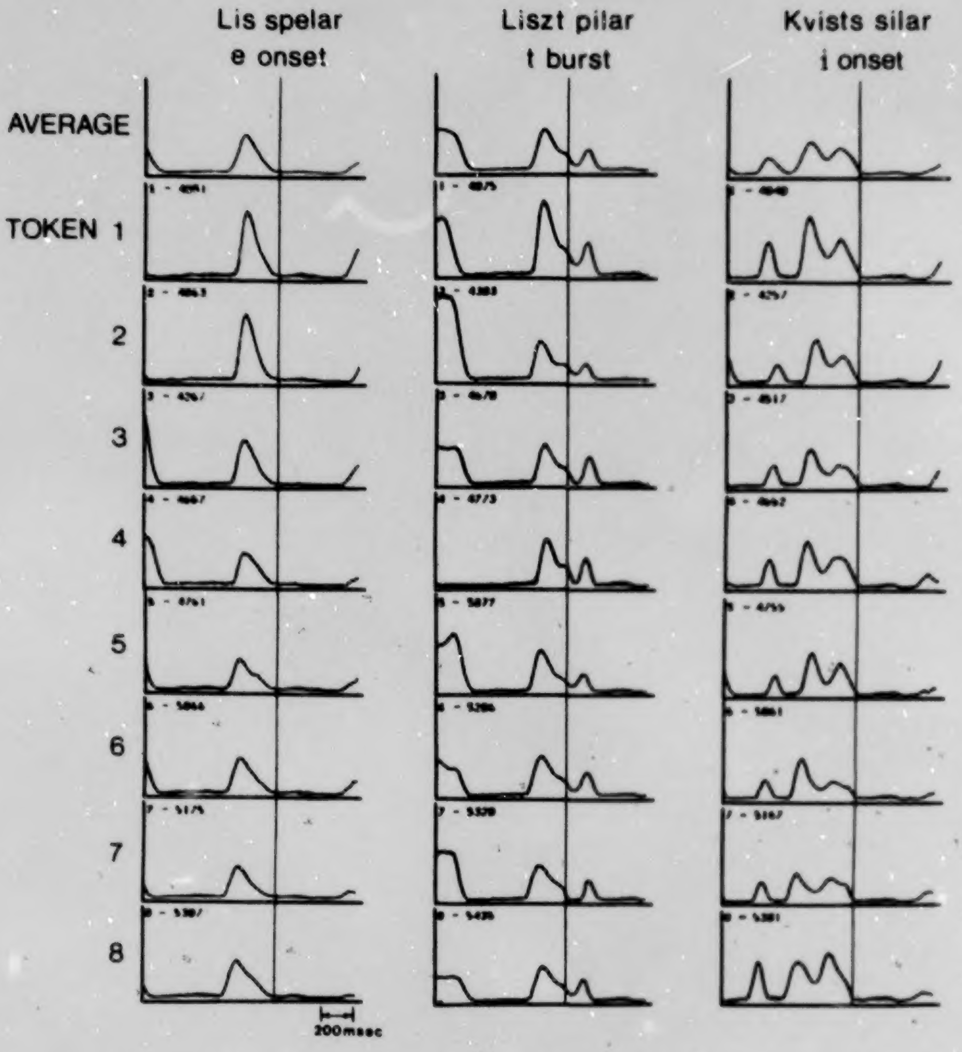


Figure 3. Average and single token transillumination signals for the utterances "Lis spelar," "Liszt pilar," and "Kvists silar." Top row shows average curves, bottom eight rows show single tokens.

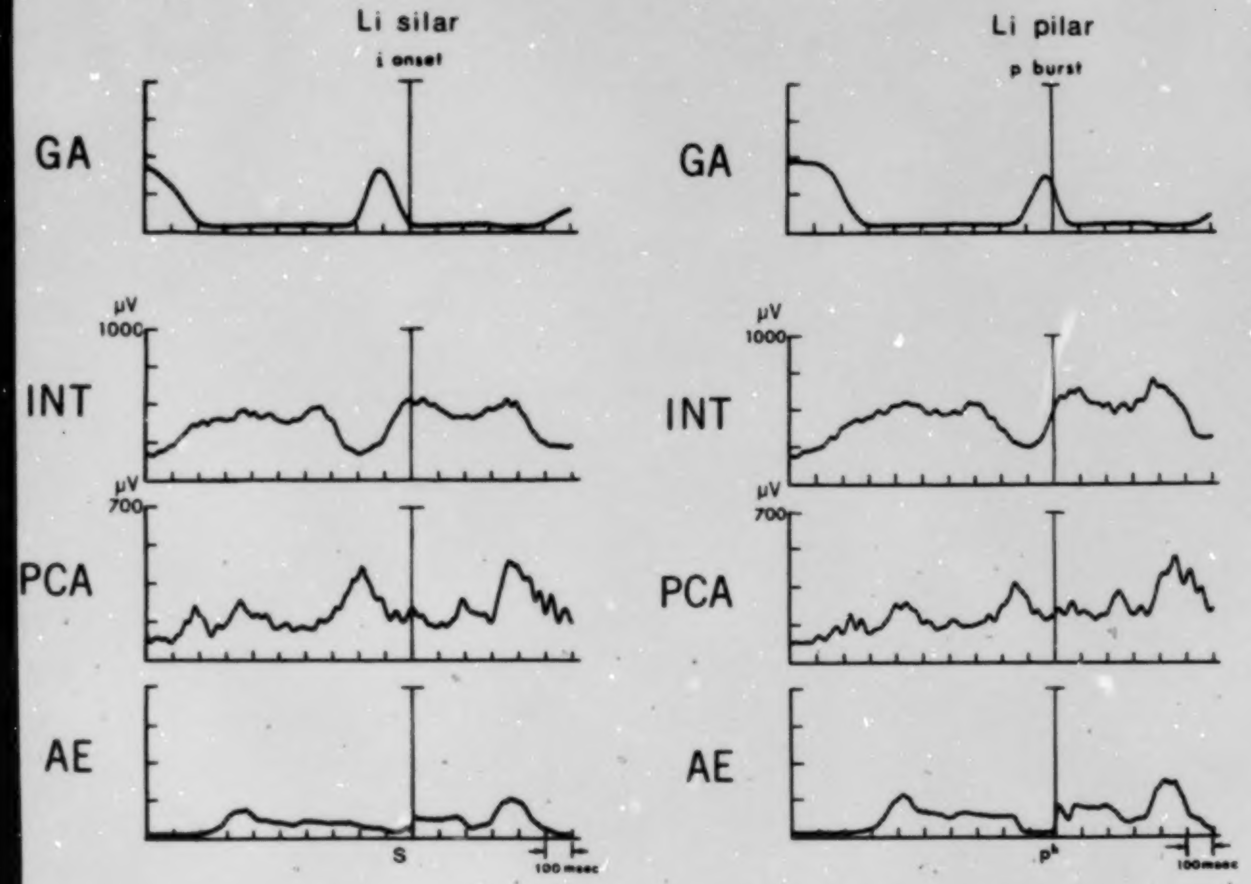


Figure 4. Average transillumination signal (GA), INT and PCA EMG records, and audio envelope (AE) for the utterances "Li silar," and "Li pilar."

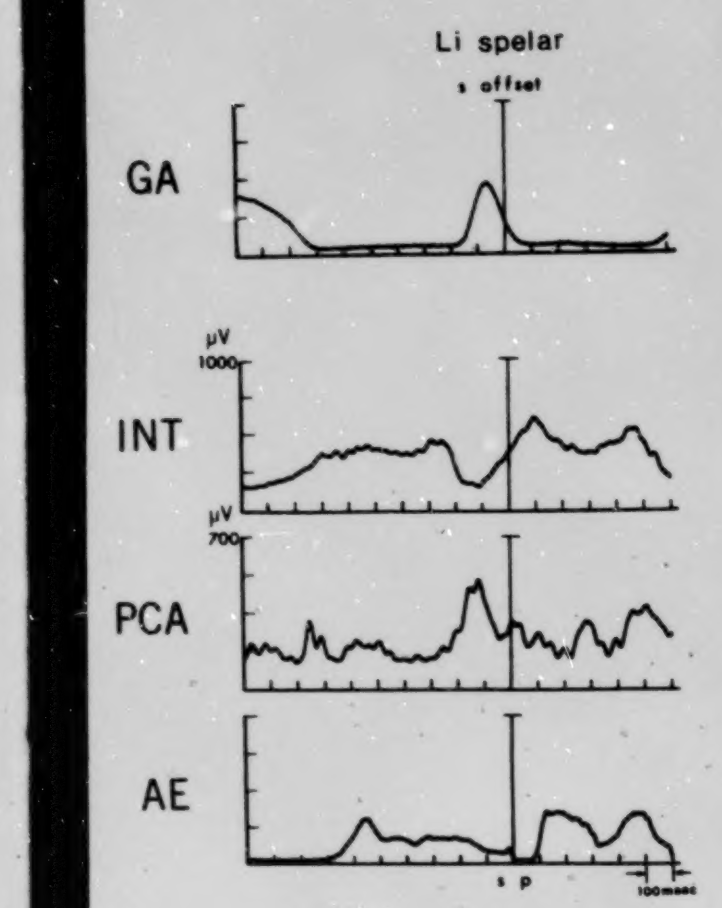


Figure 5. Glottal area, EMG, and audio signals for the utterances "Li spelar," and "Lis pilar."

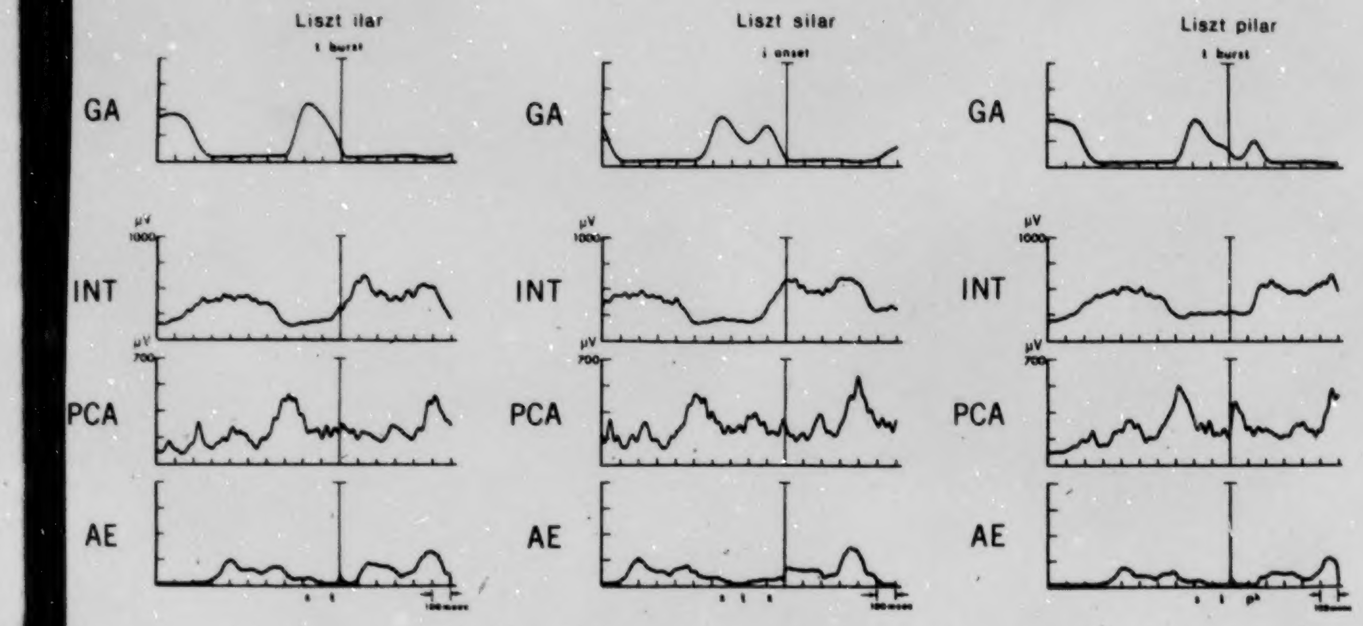


Figure 6. Glottal area, EMG, and audio signals for the utterances "Liszt ilar," "Liszt silar," and "Liszt pilar."

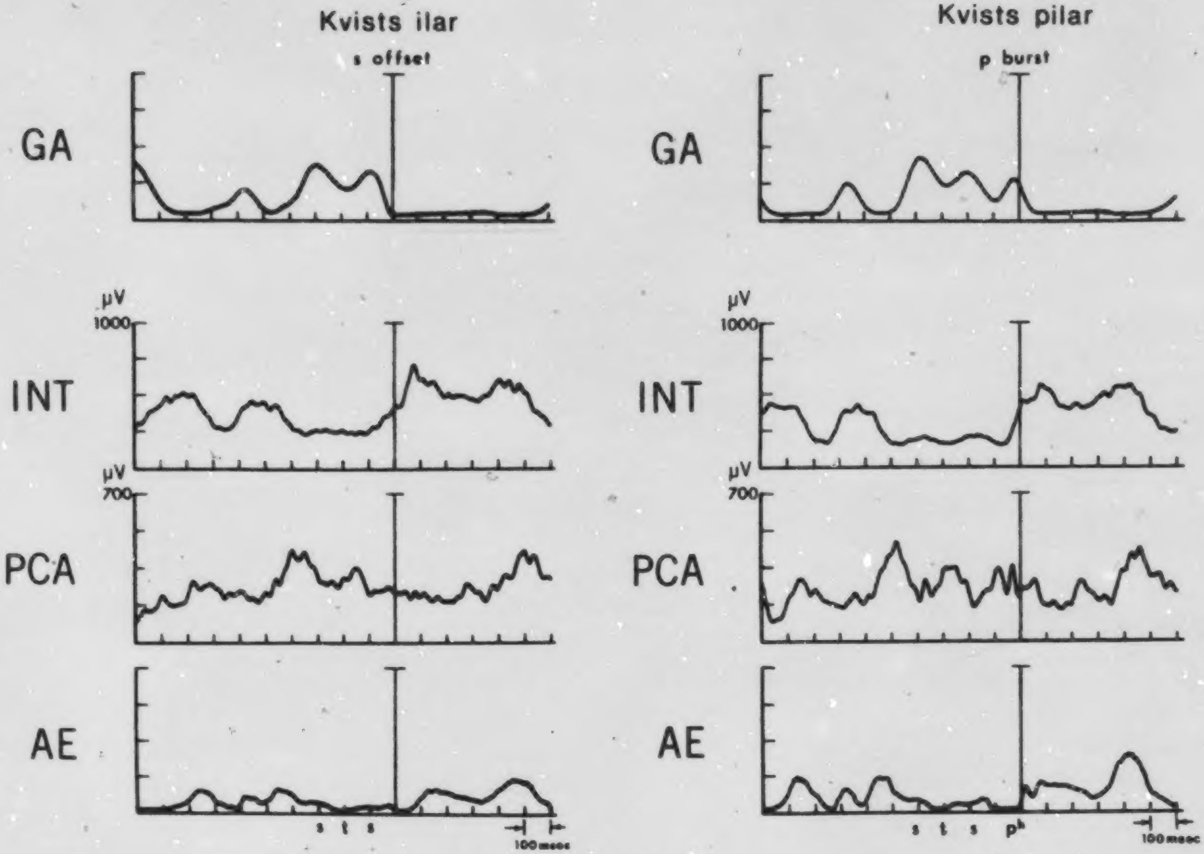


Figure 7. Glottal area, EMG, and audio signals for the utterances "Kvists ilar," and "Kvists pilar."

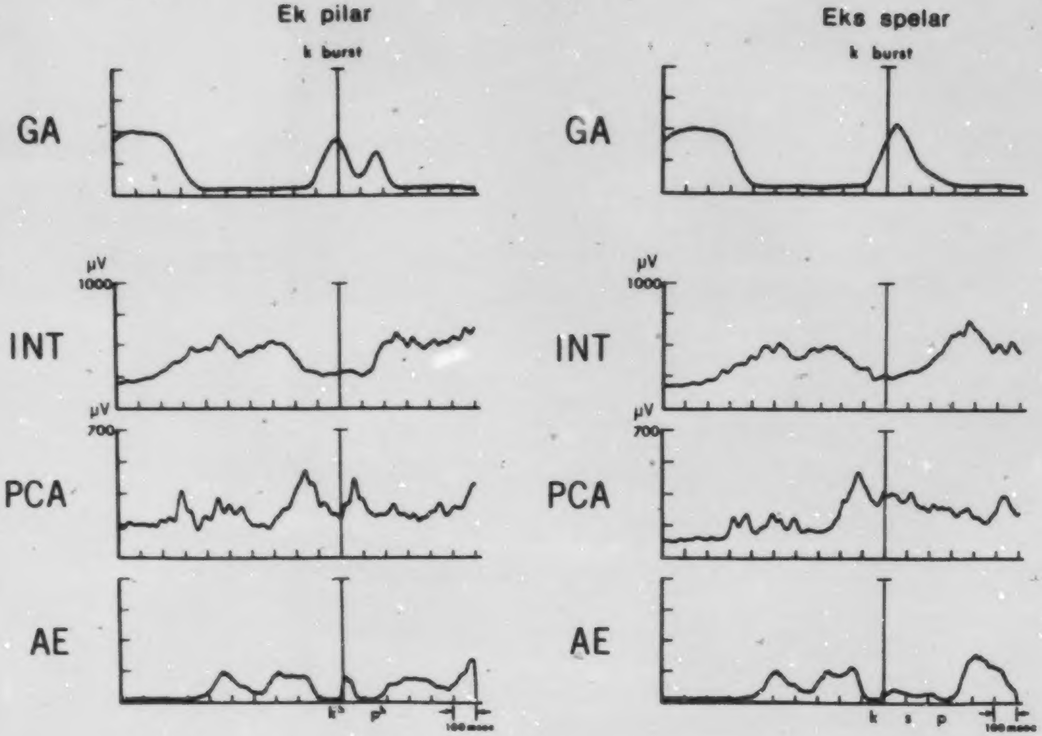


Figure 8. Glottal area, EMG, and audio signals for the utterances "Ek pilar," and "Eks spelar."

fricative.

Similar patterns are also found in clusters of voiceless obstruents. In a word initial cluster /#sp/, in Figure 5, where the stop is unaspirated according to Swedish phonology, one glottal gesture is found. This gesture is similar to the one found for a single voiceless fricative in Figure 4, with peak glottal opening occurring during the fricative. Thus, the glottis begins to close before stop implosion.

When a word boundary occurs between /s/ and /p/, Figure 5, and the stop is aspirated, certain changes occur in the pattern of glottal movements. Specifically, two consecutive gestures are found. Their shape and timing in relation to supraglottal events are similar to those found for the single obstruents above. Peak glottal opening occurs close to implosion for the fricative, and just before release of the stop. At the same time the vocal folds are adducted, though without complete glottal closure, between the two peaks of glottal opening. A minimum of glottal opening thus occurs just after stop implosion.

A similar cluster /st#/ in word final position, Figure 6, is produced with one laryngeal articulatory gesture, but its temporal course differs from that for /#sp/ in Figure 5. The glottis opens more quickly and closes more slowly so there is still some glottal opening at release of /t/. In accordance with the patterns discussed above, peak glottal opening occurs during the fricative.

In the cluster /st#s/, Figure 6, two separate gestures occur. The initial abduction is rapid, and the following adductory and abductory movements are slower. Peak glottal openings are found during the fricatives and a minimum of glottal opening just before stop release. For the cluster /st#p/ in Figure 6, the initial abduction gesture for the fricative and the gesture for the aspirated voiceless stop are similar to the ones already discussed. For the word final /t/, however, the pattern differs from that found in /st#/ in the same figure, in that there is a small extra adjustment in contrast to the overall reduction in speed of glottal adduction noted before. Thus, the glottal gesture for the word final /t/ differs according to whether a glottal stop or a voiceless aspirated stop follows the word boundary.

In the cases discussed so far, a word boundary intervened within the cluster in those instances where several consecutive articulatory gestures occurred. Even in the absence of a word boundary, multiple laryngeal articulatory movements may occur. Figure 7 presents the cluster /sts#/ with two articulatory gestures. Their relationship to oral articulations is the same as the ones noted above. The same basic pattern is also found in the cluster /sts#p/ in the same figure with three separate gestures.

Two stops follow each other with an intervening word boundary in Figure 8. Each stop is aspirated and produced with a separate laryngeal gesture. The timing of the gesture is similar for both of them, and almost identical with the gesture for a single stop in Figure 4, with peak glottal opening occurring just before release.

Figure 8 also presents the cluster /ks#sp/. Here only one laryngeal gesture occurs with peak opening during the fricative. The abduction is rather slow and related to the occurrence of maximum opening. Glottal adduction is slow, and starts before implosion of the labial stop following the fricative.

Motor control of laryngeal articulatory movements

An inspection of the combined records of glottal movements and muscular activity patterns reveals that the articulatory movements are associated with distinct activity patterns of the posterior cricoarytenoid (PCA) and the interarytenoid (INT) muscles. This holds true irrespective of the number of glottal opening and closing gestures in the clusters. The activity patterns of PCA and INT show, in general, that these two muscles are activated for glottal abduction and adduction, respectively. The present results indicate, however, that during a voiceless obstruent cluster where the glottis is open, changes in glottal opening area seem mainly controlled by the PCA. When there are more than one opening and closing gesture during the cluster, as in Figures 5, 6, 7 and 8, and the vocal folds are adducted without complete glottal closure, both the abduction and the adduction appear due to activation and inactivation of PCA, respectively. In these cases there may or may not be concomitant increased INT activity associated with the decrease in glottal opening area. Examples of increased INT activity can be seen in Figure 5 ("Lis pilar"), Figure 6 ("Liszt silar"), Figure 7 ("Kvists pilar"), and Figure 8 ("Ek pilar"). Examples of suppressed INT acitivity throughout the cluster occur in Figure 6 ("Liszt pilar") and Figure 7 ("Kvists ilar"). In all clusters, the changes in PCA activity associated with the glottal movements during the cluster are much more salient than the changes in INT. Another observation is also relevant here. For "Liszt ilar" in Figure 6 and, notably, "Eks spelar" in Figure 8, the beginning of glottal adduction, i.e., peak glottal opening, occurs well ahead of the increase in INT activity, whereas the decrease in PCA activity occurs at the appropriate time. In these cases a decrease in PCA activity thus seems more directly related to initiation of glottal adduction than INT activity per se.

The differences between the voiceless fricative and the voiceless stop in Figure 4 are partially visible in the electromyographic records. Peak PCA activity is higher for the fricative than for the stop, and the decrease in INT activity is deeper and more rapid for the fricative. Comparisons between the cluster /#sp/ in Figure 5 and the cluster /st#/ in Figure 6 reveal that the longer glottal opening and the slower adduction in the latter are accompanied by a broader peak of PCA activity.

The relationship between averaged PCA activity and glottal opening is presented in Figure 9. The plot shows the maximum glottal opening and the associated PCA activity, and also the minimum glottal opening and the concomitant PCA activity for those clusters where more than one glottal opening and closing gesture occurred. For the whole material (n=45) a correlation of +0.84 was found between glottal opening area and average PCA activity. Although such a positive relation holds for the pooled data, it is evident in Figure 9 that the same value of PCA activity can be associated with different degrees of glottal opening, and, conversely, that the same degree of glottal opening can occur with different levels of PCA activity. At the same

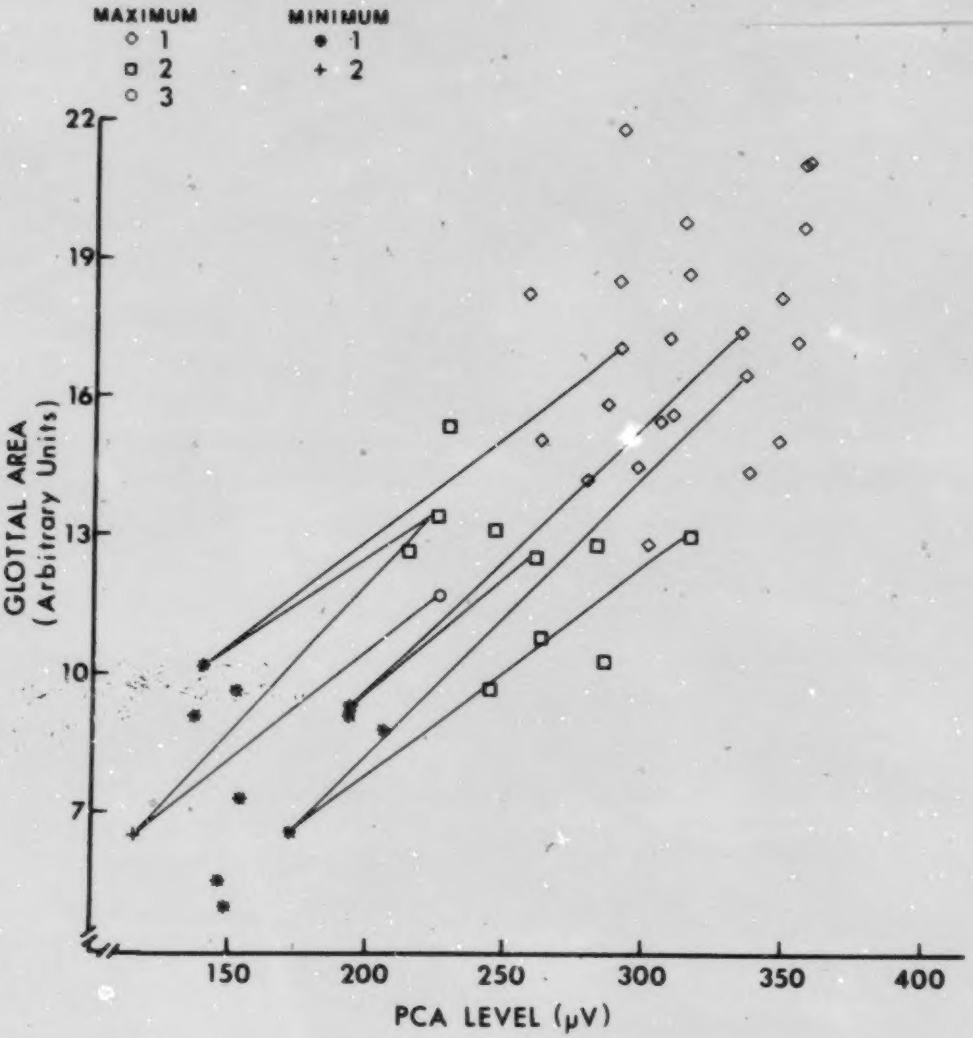


Figure 9. Average glottal opening area plotted versus corresponding average PCA activity level. Maximum 1, 2, and 3 refer to first, second and third occurrence of peak glottal opening; minimum 1 and 2 refer to intervening minima of glottal opening. Lines connect successive glottal states for three utterances.

time some regularities do exist. When several glottal openings occur, the first one is generally larger and associated with higher PCA activity. Within one and the same utterance type, the temporal changes of glottal opening area and PCA activity level are monotonically related. This is indicated by the lines in the graph connecting different data points. These lines connect successive glottal states, and the associated PCA activity level, within one and the same utterance type. For clarity of exposition only three sets of data points have been connected in this way, but a similar relationship holds true for the other utterance types as well.

DISCUSSION

The results of the present study show a high correlation between measures of glottal area variations obtained by fiberoptic filming and by transillumination. For this to hold true, it was necessary to position the phototransistor just below the cricoid cartilage. Placement of the transistor on the cricothyroid membrane made the system sensitive to vertical movements of the larynx. These movements resulted in baseline shifts related to the intonation pattern of the utterance, as well as in spuriously large glottal openings for velar sounds.

It is, at present, not possible to calibrate the transillumination system. Similarly, measurements made from fiberoptic films may contain some errors due to vertical movements of the larynx changing the distance between the glottis and the lens of the fiberscope. A better understanding of the relationship between the transillumination signal and the associated glottal opening can, in principle, be obtained by using transillumination in combination with a stereo-fiberscope (Fujimura, Baer, & Niimi, 1979).

For the present, transillumination of the larynx appears to give as accurate a record of temporal patterns of glottal movements as fiberoptic filming. Correct placement of the phototransistor is necessary, however. At the same time the transillumination technique avoids the frame-by-frame analysis and the low sampling rate of filming. The output signal is convenient for further processing, and large amounts of data can be collected and processed in a short time. If a fiberscope is used as a light source, simultaneous films can be made at regular intervals during the recording session as a further control.

The patterns of glottal movements in cluster production observed in the present study are in general agreement with those obtained in studies using similar linguistic material in English (Yoshioka, Löfqvist, & Hirose, in press) and Icelandic (Löfqvist & Yoshioka, in press). We thus have evidence from different speakers, and languages, that laryngeal activity in voiceless obstruent clusters can be organized in one or more glottal gestures, and, furthermore, that these gestures are actively controlled by muscular adjustments.

In the present material, the initiation of glottal adduction in voiceless aspirated stops occurred before stop release. In this case, no aerodynamic factors could be responsible for the initiation of the gesture. In voiceless fricatives, the closing gesture started during a period of egressive air flow, and a Bernoulli effect could in theory assist in the adduction--in particular

128

in the absence of PCA and INT activity. This does not seem likely, however, since calculations presented by Ishizaka and Matsudaira (1972) show that a negative pressure inside the glottis only occurs for small glottal openings, and the pressure is rather small. The glottal opening during voiceless fricatives is considerably larger than that during the open part of the glottal cycle in normal phonation.

It has been suggested that PCA and INT function in a reciprocal fashion (e.g., Hirose, 1976; Hirose & Ushijima, 1978). Although this is undoubtedly true in single voiceless obstruents, the material presented here suggests a slightly different picture. During a voiceless cluster, where the glottis stays open for a long period, variations in glottal area do occur. These variations can be controlled mainly by the PCA, whereas INT is suppressed during the whole cluster. At the same time, INT is clearly activated for a following vowel. The same thing is also evident in the records published by Sawashima, Hirose, and Yoshioka (1978). Their records also suggest that speakers may differ in their respective use of PCA and INT in controlling glottal opening in speech.

The notion of reciprocity maintains that the activity levels of PCA and INT are inversely related to each other. The present material points to another possibility, i.e., both muscles may be more or less suppressed.

The qualitative nature of this interpretation should be emphasized, however. In the recordings discussed above, instances of incomplete glottal adduction were accompanied by suppressed PCA activity with none, or very little, INT activity. It is, however, impossible to infer the influence on vocal fold position of a particular muscle on the basis of electromyographic recordings alone. Due to varying tension-length relationships in muscles, as well as the influence of other muscles, tissues and mechanical arrangements of laryngeal joints, too many variables are unknown here to permit a detailed modeling. Even though the change in PCA activity during the clusters is much greater than that of INT, we cannot uniquely determine their respective contributions to glottal area change. Two other caveats should also be mentioned. The suppression of INT during the clusters was comparable to that during inspiration, when INT can be assumed to be inactive. However, the noise level in the recording sets a lower limit to the amount of suppression that can be detected. Thus, we do not know if additional changes in INT activity occurred during its periods of suppression, since such changes would be masked by the background noise. Only recordings with a better signal to noise ratio can decide this question. The second caveat concerns the possible role of other adductor muscles in the absence of INT activity. Due to the technical problems of making simultaneous recordings of several channels of bioelectrical and biomechanical information, we did not attempt to record from other adductor muscles. There is nothing in the published literature to suggest that they should be generally active in the absence of INT activity. Since the function of the different adductor muscles is not very well known, we are currently making additional recordings to clarify their role in obstruent cluster production.

In Figure 9, a positive relationship is found between average glottal opening area and PCA activity. A similar relationship has been presented by Hirose (1976) and by Hirose and Ushijima (1978), and it is perhaps not a surprising finding considering the anatomical arrangement of the PCA. In view of the rather crude measures used, in particular average peak, or minimum, PCA activity, and in view of possible non-linearities in the transillumination signal, some variability in the relationship can be expected. In addition, variations in speed of glottal opening gestures and in duration of PCA activity can also account for some of the variability. In the case of the first peak of glottal opening, Maximum 1 in Figure 9, the initial condition of the glottis is the same, i.e., the glottis is in a phonatory position. For the second peak opening, Maximum 2, the initial condition is not invariant, since the data points for the preceding glottal state, Minimum 1, have different values on the y-coordinate. The three points for Maximum 2 in the utterances connected by lines in Figure 9 differ little in glottal area but more in PCA activity. At the same time, the associated preceding points, Minimum 1, also differ in these three cases. Here, similar glottal openings seem to be achieved by different amounts of PCA activity due to varying initial conditions. As can be expected, PCA activity thus seems more directly related to changes in glottal area than to glottal area per se.

The observed variations in glottal area are obviously related to various segmental properties of an utterance. The most apparent relation is that sounds requiring a high rate of air flow are produced with a separate glottal opening gesture.

It is possible that the observed incomplete glottal adduction is made to prevent excessive air flow, and waste of air during an ongoing utterance. Although there is probably some substance to such an argument in general, it is troublesome that adduction often is found during a stop closure, when no egressive air flow can occur.

The difference in laryngeal movements between stops and fricatives in Figure 4 is most likely related to different aerodynamic requirements for stop and fricative production. A rapid increase in glottal area would seem to create favorable aerodynamic conditions for the turbulent noise source during voiceless fricatives (Stevens, 1971). In stops, the timing of glottal opening during the closure is part of the mechanism controlling aspiration (Löfqvist, in press).

As a preliminary to a further discussion of laryngeal control, we can identify various theoretical positions along a continuum of "linguistic" versus "physiological" explanations of speech articulation. At one extreme is a position trying to give a direct and immediate linguistic account of every single aspect of articulatory movements. This seems to have been an assumption underlying much work on coarticulation in speech, although it has seldom been stated explicitly (cf. Daniloff & Hammarberg, 1973; Hammarberg, 1976; Bladon, 1979). The other extreme is a view that completely disregards linguistic notions, and only invokes physiological explanations. Such an approach to speech articulation has been outlined by Moll, Zimmerman, and Smith (1977). Between these extremes we can locate a number of intermediate positions that hold that linguistic aspects of an utterance are encoded in articulatory movements, but that some aspects of these movements reflect inherent characteristics of the motor system itself. The physiological position has emerged as a response to a futile search for linguistic units in articulatory movements. Although opposite in emphasis, the linguistic and the physiological position seem to share a common view of the nature of phonologic representations. Linguistic units are taken to be radically different from their articulatory realizations, and a process translating the former into the latter is required. A reconciliation of static phonological descriptions and articulatory dynamics could thus dissolve, or bridge, the linguistic/physiologic dichotomy. An outline of such a solution has been presented by Fowler, Rubin, Remez, and Turvey (in press).

Below we outline three different accounts of the present material. The first one is linguistic, the second is partly linguistic--partly physiological, and the third tries to go beyond the linguistic/physiological dichotomy by taking a different position on the question of phonological entities.

Let us first consider the cluster /#sp/ in Figure 5. The stop is unaspirated in this position, according to Swedish phonology, and the glottis begins to close before stop implosion. During stop closure there is no pressure drop across the glottis and hence no transglottal flow. Pressure above and below the glottis equalizes during the fricative and this state is maintained during the stop in the absence of special adjustments. Thus, laryngeal vibrations will not start until after stop release, when suitable aerodynamic conditions have been established. In the cluster one glottal opening and closing gesture occurs with peak glottal opening during the fricative. The glottis begins to close before stop implosion. This can be regarded as an example of normal anticipatory coarticulation. Some other aspects of glottal movements would, however, not seem to fit into this linguistic framework. Normally, non-contradictory articulatory movements are supposed to occur as soon as possible (Kozhevnikov & Chistovich, 1965; Henke, 1966). In the present case one might thus argue that the glottis should stay open during the whole cluster. Admittedly, it stays open, but to varying degrees. Within a linguistic framework, the problem is to account for these changes. One theoretical problem is that it is unclear how much deviation from a position is to be allowed before one can talk about a significant change. Studies of coarticulation have never addressed this problem, and it seems to have been tacitly assumed that any change in position would count as a significant movement. Neglecting this problem for the moment, it is possible to outline a strictly linguistic account of the observed laryngeal activity. We can thus argue that unaspirated voiceless stops after /s/ require a small glottal opening. We can also say that a linguistic boundary is associated with glottal adduction, and argue that when voiceless fricatives occur before and after the boundary, as in Figure 8, the adduction is not realized and the gestures for the two fricatives are fused into one.

A strictly physiological account is not known to us. We can, however, argue from an intermediate position, and suggest a contributing physiological factor. There is little, if any, evidence that the glottis ever opens and maintains a static open position in speech. Thus, for single voiceless obstruents, the glottis executes a "ballistic" opening and closing gesture, and in clusters one or more gestures can occur. This "cyclical" mode of glottal control may be a biologically basic phenomenon of laryngeal control, possibly related to the valving function of the larynx and its borderline position with respect to voluntary control. The cyclical activity would thus constitute the system's contribution to observed glottal movements. At the same time, these cyclical movements do not occur randomly. Rather, they are precisely coordinated with supralaryngeal events to meet the aerodynamic requirements for producing a signal with a specified acoustic structure. This would then constitute a superimposed linguistic modulation of glottal movements.

Outside the linguistic/physiological dichotomy a third account suggests itself. A glottal opening and closing gesture may be an inherent feature of different articulatory units requiring a high rate of air flow. Voiceless fricatives and voiceless stops, as well as clusters of fricative + unaspirated stop and stop + fricative, would thus have a glottal opening. The glottal abduction gesture may be intrinsically tied to movements of the upper articulators, and the oral and laryngeal component gestures of the unit executed simultaneously. Variations in glottal area during a voiceless cluster would thus follow from the temporal spacing of successive realizations of the articulatory gestures of different units.

A strict linguistic/physiological dichotomy seems wrong on theoretical grounds, in particular a requirement of direct linguistic interpretations of every single aspect of articulatory dynamics. We would thus reject the first account above. There may not be a logical incompatibility between the second and third accounts, if we leave aside the problem of properly specifying the entities involved. They do, however, differ in their emphasis on different aspects of motor control. Both share a distinct advantage over the first account in that they directly suggest further experiments in order to clarify the underlying mechanisms.

In the second case, it seems important to explore the nature of laryngeal control, in particular the amount of voluntary control that a speaker can have over glottal abduction/adduction. Scaling experiments, with and without suitable feedback, similar to those performed on velar control (Shelton, Harris, Sholes, & Dooley, 1970), and on tongue control (Porter & Lubker, 1978) can help in further clarifying this problem. Simultaneous monitoring of muscular activity and movements during such experiments can also provide information for more detailed modeling of laryngeal biomechanics.

As to the third account, suitable monitoring of oral and laryngeal articulatory movements in obstruent production could reveal the nature of their coordination. Measurements can thus be made of interarticulator timing including dynamic variables such as velocity and acceleration, and show if fixed relations exist between aspects of oral and laryngeal articulatory movements. Material presented in Löfqvist (1978) indicates that the interval from implosion to the occurrence of peak glottal opening in voiceless fricatives may remain almost constant, irrespective of variations in overall fricative duration.

REFERENCES

Basmajian, J., & Stecko, G. A new bipolar indwelling electrode for electromyography. Journal of Applied Physiology, 1962, 17, 849.

Bladon, R. A. Motor control of coarticulation: Linguistic considerations. In Proceedings of the Ninth International Congress of Phonetic Sciences, Vol. 2, 325-331, 1979.

Coleman, R., & Wendahl, R. On the validity of laryngeal photosensor monitoring. *Journal of the Acoustical Society of America*, 1968, *44*, 1733-1735.

Daniloff, R., & Hammarberg, R. On defining coarticulation. *Journal of Phonetics*, 1973, *1*, 239-248.

Fowler, C. A., Rubin, P., Remez, R., & Turvey, M. Implications for speech production of a general theory of action. In B. Butterworth (Ed.), *Speech production*. New York: Academic Press, in press.

Frøkjaer-Jensen, B. Comparison between a Fabre glottograph and a photoelectric glottograph. *Annual Report* (Institute of Phonetics, University of Copenhagen), 1968, *3*, 9-16.

Fujimura, O., Baer, T., & Niimi, S. A stereo-fiberscope with magnetic interlens bridge for laryngeal observation. *Journal of the Acoustical Society of America*, 1979, *65*, 478-480.

Fujimura, O., & Sawashima, M. Consonant sequences and laryngeal control. *Annual Bulletin* (Research Institute of Logopedics and Phoniatrics, University of Tokyo), 1971, *5*, 1-6.

Hammarberg, R. The metaphysics of coarticulation. *Journal of Phonetics*, 1976, *4*, 353-363.

Harden, R. J. Comparison of glottal area changes as measured from ultrahigh-speed photographs and photoelectric glottographs. *Journal of Speech and Hearing Research*, 1975, *18*, 728-738.

Henke, W. L. *Dynamic articulatory model of speech production using computer simulation*. Unpublished doctoral dissertation, Massachusetts Institute of Technology, 1966.

Hirano, M., & Ohala, J. Use of hooked-wire electrodes for electromyography of intrinsic laryngeal muscles. *Journal of Speech and Hearing Research*, 1969, *12*, 362-373.

Hirose, H. Posterior cricoarytenoid as a speech muscle. *Annals of Otology, Rhinology and Laryngology*, 1976, *85*, 334-342.

Hirose, H., & Gay, T. The activity of the intrinsic laryngeal muscles in voicing control. An electromyographic study. *Phonetica*, 1972, *25*, 140-164.

Hirose, H., & Ushijima, T. Laryngeal control for voicing distinction in Japanese consonant production. *Phonetica*, 1978, *35*, 1-10.

Hirose, H., Yoshioka, H., & Niimi, S. A cross language study of laryngeal adjustment in consonant production. *Annual Bulletin* (Research Institute of Logopedics and Phoniatrics, University of Tokyo), 1978, *12*, 61-71.

Ishizaka, H., & Matsudaira, M. Fluid mechanical considerations of vocal cord vibration. *Monograph No. 8* (Speech Communications Research Laboratory, Santa Barbara), 1972.

Kewley-Port, D. EMG signal processing for speech research. *Haskins Laboratories Status Report on Speech Research*, 1977, *SR-50*, 123-146.

Kitzing, P. Methode zur kombinierten photo- und elektroglottographischen Registrierung von Stimmlippenschwingungen. *Folia Phoniatrica*, 1977, *29*, 249-260.

Kitzing, P., & Löfqvist, A. Clinical application of combined electro- and photoglottography. In N. Buch (Ed.), *Proceedings of the 17th International Congress of Logopedics and Phoniatrics*. Copenhagen: Special-Paedagogisk Forlag, 1978.

Köster, J-P., & Smith, S. Zur Interpretation elektrischer und photoelektrischer Glottogramme. *Folia Phoniatrica*, 1970, *22*, 92-99.

Kozhevnikov, V., & Chistovich, L. *Speech: Articulation and perception*. Moscow-Leningrad, 1965 (English translation: J.P.R.S., Washington, D.C. No. JPRS 30543).

Löfqvist, A. Artikulatorisk programmering - några data kring svenska obstruentkombinationer. Nysvenska Studier, 1977, 57, 95-104.

Löfqvist, A. Laryngeal articulation and junctures in the production of Swedish obstruent sequences. In E. Gårding, R. Bruce, & R. Bannert (Eds.), Nordic prosody. (Travaux de l'Institut de Linguistique de Lund XIII.) Lund: Department of Linguistics, 1978.

Löfqvist, A. Interarticulator programming in stop production. To appear in Journal of Phonetics, in press.

Löfqvist, A., & Yoshioka, H. Laryngeal activity in Icelandic obstruent production. Paper prepared for the Fourth International Conference of Nordic and General Linguistics, in press.

Moll, K. L., Zimmerman, G., & Smith, A. The study of speech production as a human neuromotor system. In M. Sawashima & F. S. Cooper (Eds.), Dynamic aspects of speech production. Tokyo: University of Tokyo Press, 1977.

Pétursson, M. Jointure au niveau glottal. Phonetica, 1978, 35, 65-85.

Porter, R., Jr., & Lubker, J. Some interesting things you can do with your tongue. Journal of the Acoustical Society of America, 1978, 63, S52. (Abstract)

Sawashima, M., Hirose, H., & Yoshioka, H. Abductor (PCA) and adductor (INT) muscles of the larynx in voiceless sound production. Annual Bulletin (Research Institute of Logopedics and Phoniatrics, University of Tokyo), 1978, 12, 53-60.

Shelton, R., Harris, K. S., Sholes, G., & Dooley, P. Study of nonspeech voluntary palate movements by scaling and electromyographic techniques. In J. F. Bosma (Ed.), Second symposium on oral sensation and perception. Springfield, Ill.: Thomas, 1970.

Sonesson, B. On the anatomy and vibratory pattern of the human vocal folds. Acta Oto-laryngologica, 1960, (Supplementum 156).

Stevens, K. Airflow and turbulent noise for fricative and stop consonants: Static considerations. Journal of the Acoustical Society of America, 1971, 50, 1180-1192.

Yoshioka, H., Löfqvist, A., & Hirose, H. Laryngeal adjustments in the production of consonant clusters and geminates in American English. Haskins Laboratories Status Report on Speech Research, SR-59/60, this issue.

134

LARYNGEAL ADJUSTMENTS IN THE PRODUCTION OF CONSONANT CLUSTERS AND GEMINATES IN AMERICAN ENGLISH*

Hirohide Yoshioka+, Anders Löfqvist++ and Hajime Hirose+++

Abstract. The glottal opening gesture and its timing control in various sequences of voiceless obstruents were investigated by the combined techniques of electromyography, photo-electric glottography and fiberoptic endoscopy. Two distinct peaks in the abductor muscle (PCA) activity curve were found for the /sk/ sequence when a word boundary intervened and, consequently, the /k/ was aspirated, but only one peak for the same sequence without the boundary. For the geminate combination /ss/ or /kk/, which was pronounced as a single obstruent accompanied by long duration of frication or closure, the abductor activity pattern, as well as the corresponding glottal opening curve, was characterized by one single peak surrounded by gradual slopes, although a word boundary intervened within the geminate. In the case where word initial aspirated /k/ was preceded by word final cluster /sk/ or /ks/, however, the abductor muscle showed a bimodal activity pattern during the whole voiceless sequence. Furthermore, the temporal patterns of the glottal opening movement registered by photo-electric glottography and fiberoptic endoscopy revealed that the first opening gesture was at its maximum during the fricative /s/ segment, and that the second one reached its maximum around the burst of the stop. The results gathered at both electromyographic and movement levels in this experiment, including those mentioned above, suggest that each voiceless obstruent specified by aspiration or frication noise tends to require a single separate opening gesture, while an unaspirated stop in a voiceless environment can be produced within the opening gesture attributed to an adjacent aspirated stop or fricative. Such an independent opening gesture of the glottis for the production of voiceless aspirated stops or voiceless fricatives can be interpreted as assuring the aerodynamic requirement for turbulent noise production during the aspirated stop or fricative segment.

---

*A version of this paper was presented at the 97th Meeting of the Acoustical Society of America in Cambridge, Massachusetts, 11-15 June 1979.

+Also University of Tokyo, Tokyo, Japan.

++Also Lund University, Lund, Sweden.

+++University of Tokyo, Tokyo, Japan.

Acknowledgment: This work was supported in part by NINCDS Grants NS-13617 and NS-13870, and BRS Grant RR-05596 to Haskins Laboratories. The authors are grateful to Thomas Baer who participated as the subject.

[HASKINS LABORATORIES: Status Report on Speech Research SR-59/60 (1979)]

INTRODUCTION

It has been universally recognized that the larynx plays the major role in accomplishing the phonemic distinction of voicing. Although there still remain some arguments on the details of the transition from voicing to no voicing, and vice versa--as well as on the participation of other articulators--the approximation and separation of the vocal folds, in particular, are considered crucial conditons among other physical or aerodynamic factors determining the initiation, maintenance and cessation of the vibrations (e.g., Fant & Scully 1977). Many studies, using photo-electric glottography and fiberoptics, have confirmed that the precise degree and timing of this glottal opening and closing gesture are critically linked to the manifestion of the distinctive features not only of voicing but also of aspiration in several languages (Frøkjaer-Jensen, Ludvigsen, & Rischel, 1971; Kagaya, 1974; Dixit, 1975; Kagaya & Hirose, 1975; Iwata & Hirose, 1976).

Furthermore, the phonetic variation of aspiration in stop consonant production in English and Swedish has been shown to be based on different time courses of the opening gesture of the glottis (Lisker, Abramson, Cooper, & Schvey, 1969; Sawashima, 1970; Lindqvist, 1972; Löfqvist, 1976). As for non-distinctive variations in voicing, the vowel devoicing phenomenon in Japanese, for example, has also been demonstrated to be accompanied by an open glottis, which is chiefly responsible for this particular unvoiced allophone (Sawashima, 1971). Although the /h/ voicing phenomenon in Japanese, another type of phonetic variation in voicing, is not well explained in terms of only glottal aperture (Yoshioka, 1979), the dimension of glottal opening and closing has, at least in most voiced versus voiceless pairs, proved to be substantially correlated with the quasi-periodical excitations at the glottal level.

The electromyographic work of the past decade has further confirmed that the degree and timing of the glottal aperture are controlled, at least in gross terms, by reciprocal activity patterns of the abductor and adductor muscle groups of the larynx (Hirose & Gay 1972). Although functional differences among adductor muscles have not been well investigated in relation to glottal movements, the activity pattern of the posterior cricoarytenoid muscle, considered to be the sole abductor, has been shown to be most critical for determination of the glottal aperture in general (Hirose, 1976). Thus, combined experiments registering both electromyographic and movement parameters have revealed that the contrast of voiced versus voiceless as well as that of aspirated versus unaspirated is accounted for in terms of the underlying neuromuscular control of these muscles, except for the /h/ voicing phenomenon in Japanese, mentioned above.

These investigations, however, have dealt mainly with simple speech materials such as the alternating phoneme sequences, /CVCV/, /CVCVC/ or /VCV/ (C=voiceless consonant, V=vowel). In other words, such studies may be directed towards the understanding of laryngeal control in voicing distinctions and/or variations under circumstances of minimal mutual interaction between adjacent phones with respect to voicing. Since vowels are usually voiced and most of the consonants in many languages seem to be clearly specified by either a voiced (lax) or a voiceless (tense) feature, and indeed binary in this sense (e.g., Jakobson, Fant, & Halle, 1951), coarticulatory phenomena at the glottal level appear to be less likely, as far as the above-mentioned contexts are concerned. In this connection, the study of vowel devoicing can be interpreted as aiming at the exceptional cases where such a type of laryngeal coarticulation does occur. In contrast, many other coarticulation studies, most of which deal with supralaryngeal articulations, have focused on observing coarticulation during sounds specified as "neutral" with regard to features such as lip rounding and nasalization.

In addition to examining laryngeal coarticulation in regularly alternating phoneme combinations, it is also interesting to find out how phoneme sequences homogeneous with respect to voicing are organized at the level of the glottis in terms of their neurophysiological correlates. The present paper is intended to clarify the temporal change of the glottal opening gesture and its neuromuscular control, specifically during the production of clusters of voiceless obstruents in American English. This study may provide insight into coarticulatory phenomena at the level of laryngeal adjustments as well as additional information on the biomechanical properties of the opening and closing movements of the vocal folds from a kinesiological viewpoint. Another purpose of the study is to explore the phonetic effect, if any, of word boundaries on laryngeal articulation, since some voiceless phonemes and voiceless sequences in American English may occur in a variety of linguistic situations, i.e., they may be preceded, followed or interrupted by a word boundary.

METHOD AND PROCEDURE

The experiment was conducted in two parts; one is an electromyographic (EMG) study of the larynx and the other is a movement study using the combined techniques of photo-electric glottography and fiberoptic viewing of the glottis.

The EMG data were obtained using bipolar hooked-wire electrode techniques (Basmajian & Stecko, 1962; Hirano & Ohala, 1969). The electrodes, consisting of a pair of platinum-tungsten alloy wires (50 microns in diameter with isonel coating), were inserted perorally into the posterior cricoarytenoid muscle (PCA) under indirect laryngoscopy with the aid of a specially designed curved probe (Hirose, Gay, Strome, & Sawashima, 1971). Before the insertion, topical anaesthetic was applied to the mucous membrane of the hypopharynx using a small amount of 4% lidocaine spray (Xylocaine). The interference voltages of the EMG signals were recorded on an FM multichannel data recorder in parallel with the acoustic signal. The action potentials, then, were fed into a digital computer system and sampled at a rate of 200/sec, after being rectified and integrated over a 5-msec time window, for further processing to obtain the muscle activity patterns for ensemble-averaged tokens (Kewley-Port, 1977). The figures to be presented in this paper represent activity patterns aligned with reference to particular acoustic events, and smoothed with a time constant of 35 msec, before ensemble-averaging.

For the movement data, the glottal view through a flexible laryngeal fiberscope (Olympus VF-O type, 4.5 mm in outer diameter) was photographed with a cine camera at a rate of 60 frames/sec. Both the audio signal and the synchronization signal were registered on the FM recorder tape to identify each frame. Then, frame by frame analyses were made with the aid of a mini computer to calculate, on each frame, the distance between the vocal processes. [The distance is considered one of the indicators of glottal width (Sawashima & Hirose, 1968; Sawashima, 1976).]

A cold DC light source (OLYMPUS CLS), providing illumination of the upper glottal area, also served as the light source for the photo-electric glottography. The amount of light passing through the glottis was sensed by a photo-transistor (Philips BPX 81) placed on the neck just below the lower edge of the cricoid cartilage. The electrical output was also recorded on another channel of the FM tape. These three signals were sampled at 200/sec and processed in the digital system. They will be shown with a 5 msec integration time constant.

Table 1. Test utterance types.

1. I may aid	/ # /	9. My ace aids	/ s# /	17. He makes aid	/ks# /
2. I may sale	/ #s /	10. My ace sales	/ s#s /	18. He makes sale	/ks#s /
3. I may cave	/ #k /	11. My ace caves	/ s#k /	19. He makes cave	/ks#k /
4. I may scale	/ #sk/	12. My ace scales	/ s#sk/	20. He makes scale	/ks#sk/
5. I make aid	/ k# /	13. I mask aid	/sk# /	21. He masks aid	/sks# /
6. I make sale	/ k#s /	14. I mask sale	/sk#s /	22. He masks sale	/sks#s /
7. I make cave	/ k#k /	15. I mask cave	/sk#k /	23. He masks cave	/sks#k /
8. I make scale	/ k#sk/	16. I mask scale	/sk#sk/	24. He masks scale	/sks#sk/

A native male speaker of standard American English served as the subject. Among the possible voiceless phoneme sequences in this language, the combination of /s/ and /k/ is optimum in forming the greatest possible number of meaningful contexts. Therefore, as is shown in Table 1, "sentences" containing the phonemes /s/ and /k/ in many combinations were selected for the test utterances. The abbreviated phonemic transcriptions indicate the types of clusters with which the experiment is concerned. In the first EMG session, the subject was asked to produce the 24 utterance types, 12 times each, in random order. For the movement study, simultaneous recordings of photo-electric output and fiberoptic cine film were made during the first two repetitions of each utterance type, followed by 12 additional recordings of only the photo-electric signal. During the session, the glottal image was constantly monitored through the fiberoptic view finder. Although no particular instruction was given to the subject about the vocal intensity or the speaking rate, a gross survey of audio waveforms and acoustic envelopes revealed that the intra-session variability for each utterance type is comparable with that across sessions.

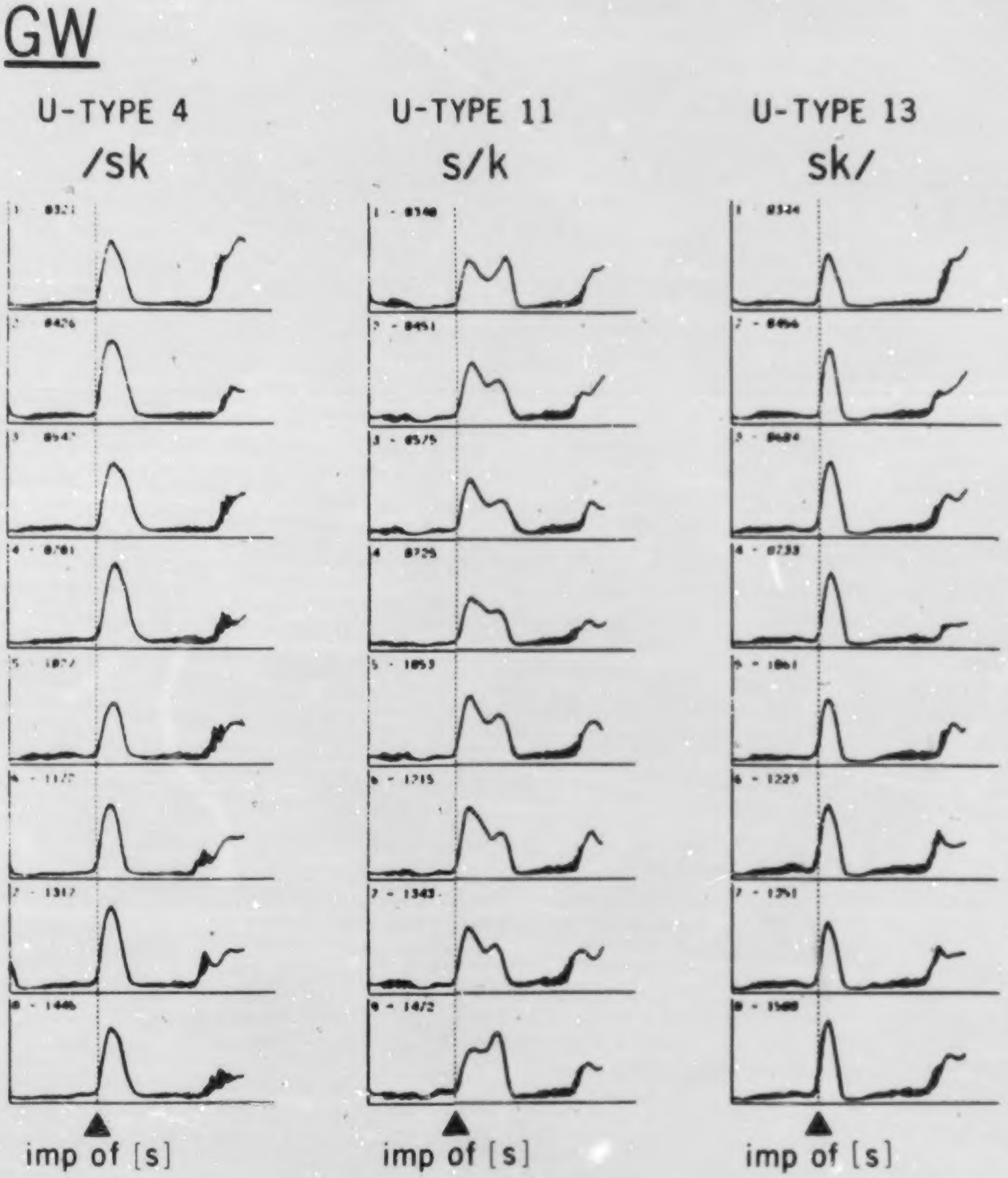
RESULTS

Figure 1 contains the glottographic patterns for the first eight tokens out of 14 productions in three utterance types, where the place of the word boundary varies in the same /sk/ sequence. In each graph, the vertical dotted line on the time axis corresponds to the implosion of [s] segment, which served as the line-up point for the sampling and averaging. More specifically, the acoustic reference was determined by identifying the voicing offset for the preceding vowel in each audio waveform of each utterance type. An overall survey of this figure reveals that, although there are some variations within each utterance type--particularly in the peak values for type 11--the glottograms for type 4 and type 13 show one single opening gesture of the glottis for the voiceless segments, while those for type 11 clearly demonstrate two separate opening gestures for the same sequences of phonemes.

In order to illustrate the corresponding activity patterns of the abductor muscle (PCA), which is presumably most responsible for the glottal aperture, Figure 2 shows its averaged electromyographic curves for the same three utterance types. In addition, for each utterance type, a representative plot of glottal width as a function of time drawn from fiberoptic data (GWf) is included along with the glottographic curve recorded simultaneously by transillumination (GWt). Also, for both electromyographic and glottal width data, the corresponding audio envelopes (AE), aligned with reference to the implosion of [s], are shown for gross segmentation of the acoustic events.[1] These curves further confirm that the difference in the number of peak openings among the productions of the /sk/ combination is inherently linked to underlying difference in the activity pattern of the abductor muscle with some time lead. That is, two distinct peaks in the PCA activity curve are found for type 11, where a word boundary intervened within the /sk/ sequence, but there is only one peak for the utterance types without the boundary. Note that the PCA activity pattern for the /s#k/ sequence in utterance type 11 shows complete relaxation down to the noise level, preceded by a peak around the line-up and followed by reactivation. The extra small peaks found in the PCA curves for utterance types 11 and 13 correspond to the release of the glottal attack for word initial vowels in the frame sentences, [?e] in "My ace caves" and [?e] in "I mask aid", respectively (Hirose & Gay, 1973).

A closer comparison of the single acoustic signals at the bottom, with the corresponding time courses of the glottal width in the middle, reveals that the peaks in glottal opening curve are always reached during the fricative segment [s] or the aspirated stop [k^h] if one occurs. No specific opening gesture is detectable for the unaspirated stop [k] in utterance types 4 and 13. Rather, the glottal articulation during this particular allophone seems to be merely a continuation of the closing phase of the glottal gesture for the preceding [s] segment, since the curves are almost symmetrical with regard to their peak timing. The peak glottal openings for utterance types 4 and 13 are quite comparable. The velocity of the glottal movement appears to be a little faster for the word-final /sk/ cluster than for word-initial /sk/ sequence in both opening and closing phases, since the frication duration is usually shorter in word-final position. In contrast, neither of the maximum values in the glottal width curve for utterance type 11 is as high as those for the other two utterance types. Taken together, the fact that both [s] and

Figure 1: Glottographic patterns for first 8 productions of three utterance types containing /sk/ sequence in various contexts.



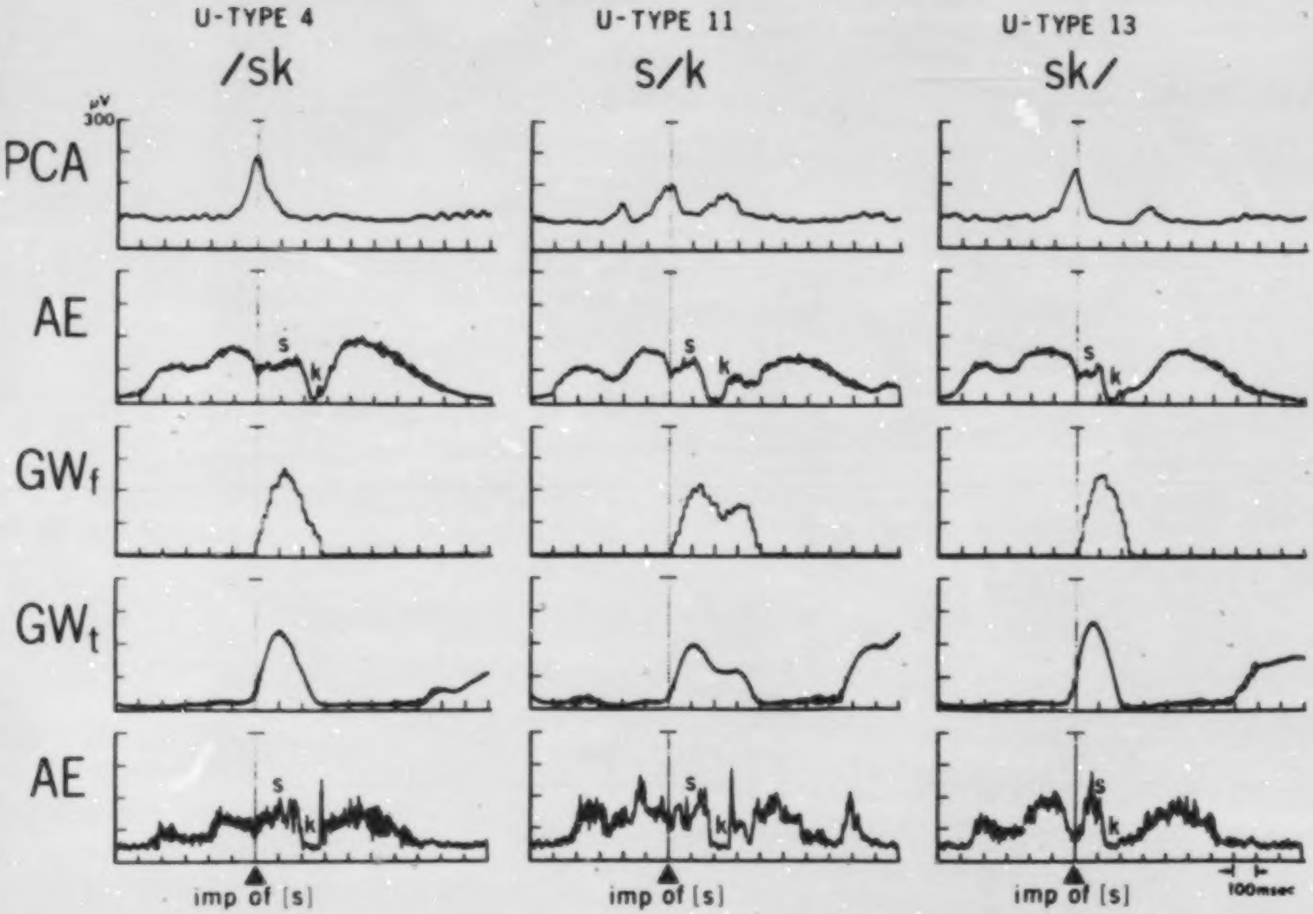


Figure 2: Averaged electromyograms of PCA, averaged audio envelopes, representative plots of glottal width using fiberoptics, corresponding glottograms and audio envelopes for same three utterance types as in Figure 1.

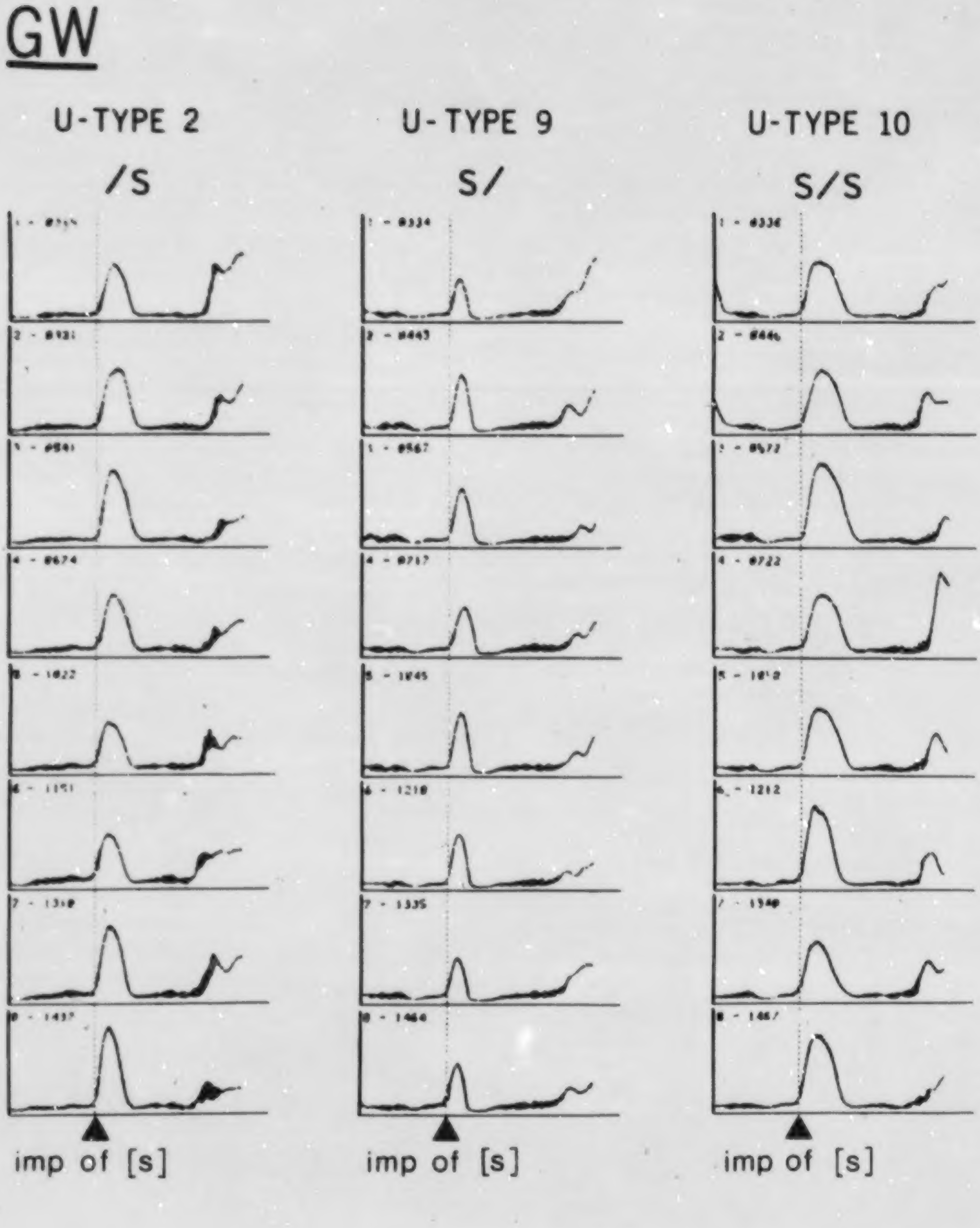
[k^h] require certain amounts of opening also means that such a temporary closing movement (shown as a dip between peaks in type 11) should not be interpreted simply as the presence of a prolonged pause at the word boundary, but as controlled narrowing in this particular context.

Figures 3 and 4 compare the productions of the phoneme /s/ in various contexts including a geminate combination. It should be mentioned here that, as is implied by the acoustic signals, the geminated sound /s#s/ was produced with prolonged continuous frication noise. The averaged abductor activity curves (shown in Figure 4) as well as the glottographic patterns for the first eight tokens (shown in Figure 3) among these three utterance types are all characterized by one single opening peak, regardless of word boundary position. Besides the additional small peaks in the PCA curves, which correlate with the glottal attack of vowels, the detailed patterns of the curves differ in several aspects, however. The maximum opening for word final /s/ is significantly smaller than that for word-initial /s/, and the frication period is shorter in final position. Consequently, the velocity of the opening as well as the closing phases does not seem significantly different in either position. Although the movement curves for these single cases of /s/ are both nearly symmetrical around their peaks, the curve for the geminated one /s#s/ appears to be specified by slower velocity of its closing phase. In other words, the glottal opening during the geminate sequence reaches the maximum as quickly as those during the single ones, but the width decreases slowly until the end of the prolonged frication.

Figures 5 and 6 contain the production of the geminate combination /k#k/ in contrast to the corresponding single voiceless stops. The acoustic signals reveal that the geminated sound was uttered with a longer duration of the closure period followed by a certain degree of aspiration, comparable to that for a single aspirated stop [k^h]. The curve of the abductor activity pattern, as well as those of glottal opening for the geminate, also appears to be characterized by one single peak similar to that for the word initial aspirated [k^h], although a word boundary intervenes within the geminate. In addition, the stop burst, indicated by arrows in the graphs, shows that, at least in this subject, the glottal opening is at its maximum during the aspiration period for the single stop, while it peaks before, or around, the burst for the geminate cognate. In contrast, the word-final /k/ is completely different. In the glottographic figures, the lower and upper pointed triangles correspond to the implosion of the silence and its release for the glottal attack of the following vowel, respectively. The data clearly demonstrate that the word-final stop /k/ was actually produced with a negligibly small opening gesture of the glottis, presumably due to glottalization in this particular position.

Figures 7 and 8 further compare three phone combinations similar to each other. Although the number of peak openings is not always easy to count, there is a general tendency for each type of voiceless cluster in these utterances to be produced with essentially two separate peaks of the opening gesture at both the electromyographic and the movement levels. Moreover, a gross acoustic segmentation, by inspection of the acoustic envelope, makes it possible to identify the affiliation of each separate opening gesture. The word-initial /s/ and /k/ appear to be produced with a single opening gesture, while the word-final /sk/ and /ks/ are pronounced within another separate

Figure 3: Glottographic patterns for first 8 productions of three utterance types containing fricative /s/ in various contexts.



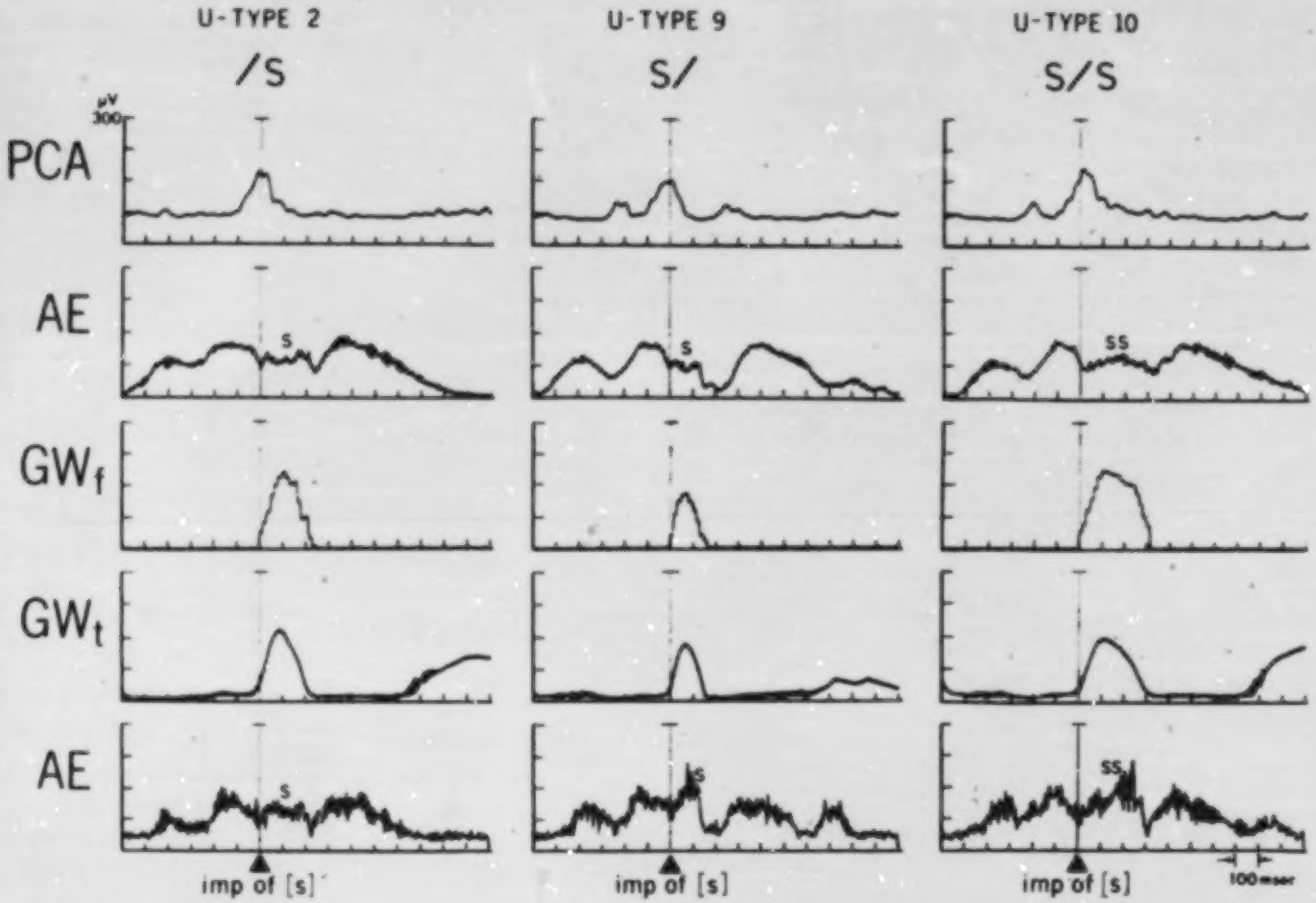
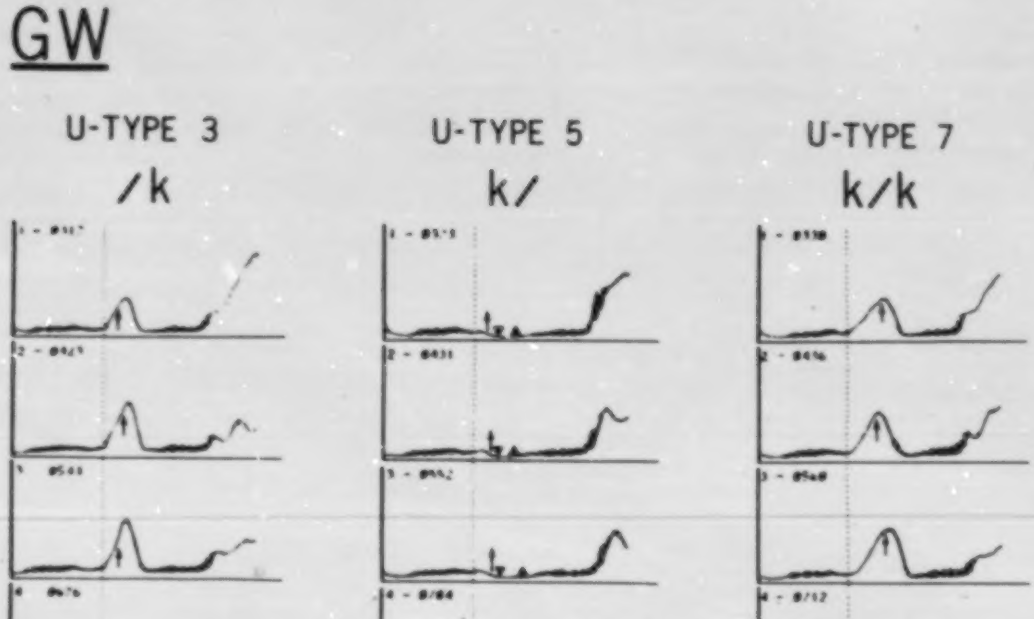


Figure 4: Averaged electromyograms of PCA, averaged audio envelopes, representative plots of glottal width using fiberoptics, corresponding glottograms and audio envelopes for same three utterance types as in Figure 3.

Figure 5: Glottographic patterns for first 8 productions of three utterance types containing stop /k/ in various contexts.



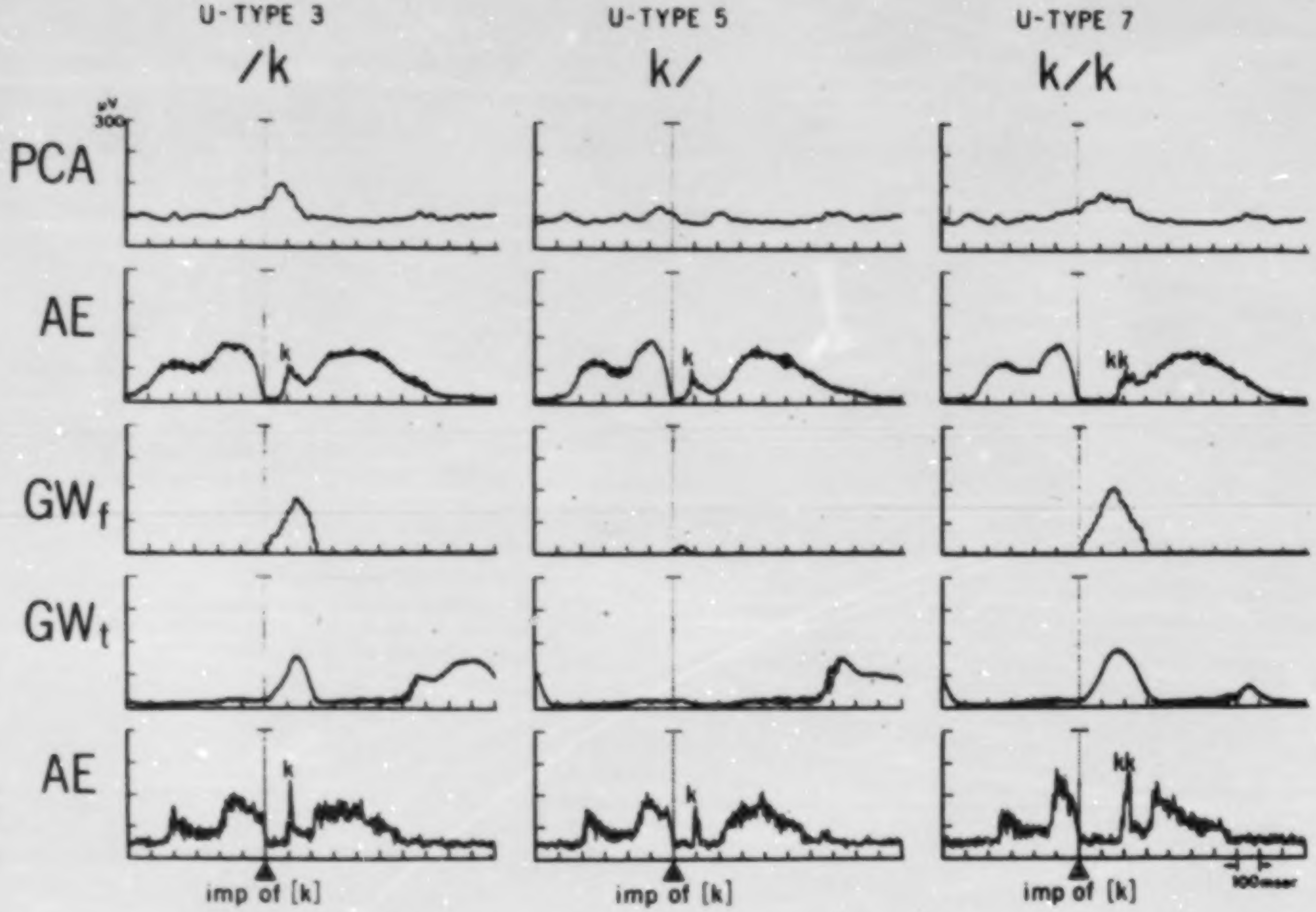


Figure 6: Averaged electromyograms of PCA, averaged audio envelopes, representative plots of glottal width using fiberoptics, corresponding glottograms and audio envelopes for same three utterance types as in Figure 5.

Figure 7: Glottographic patterns for first 8 productions of three utterance types containing various combinations of three voiceless phones.

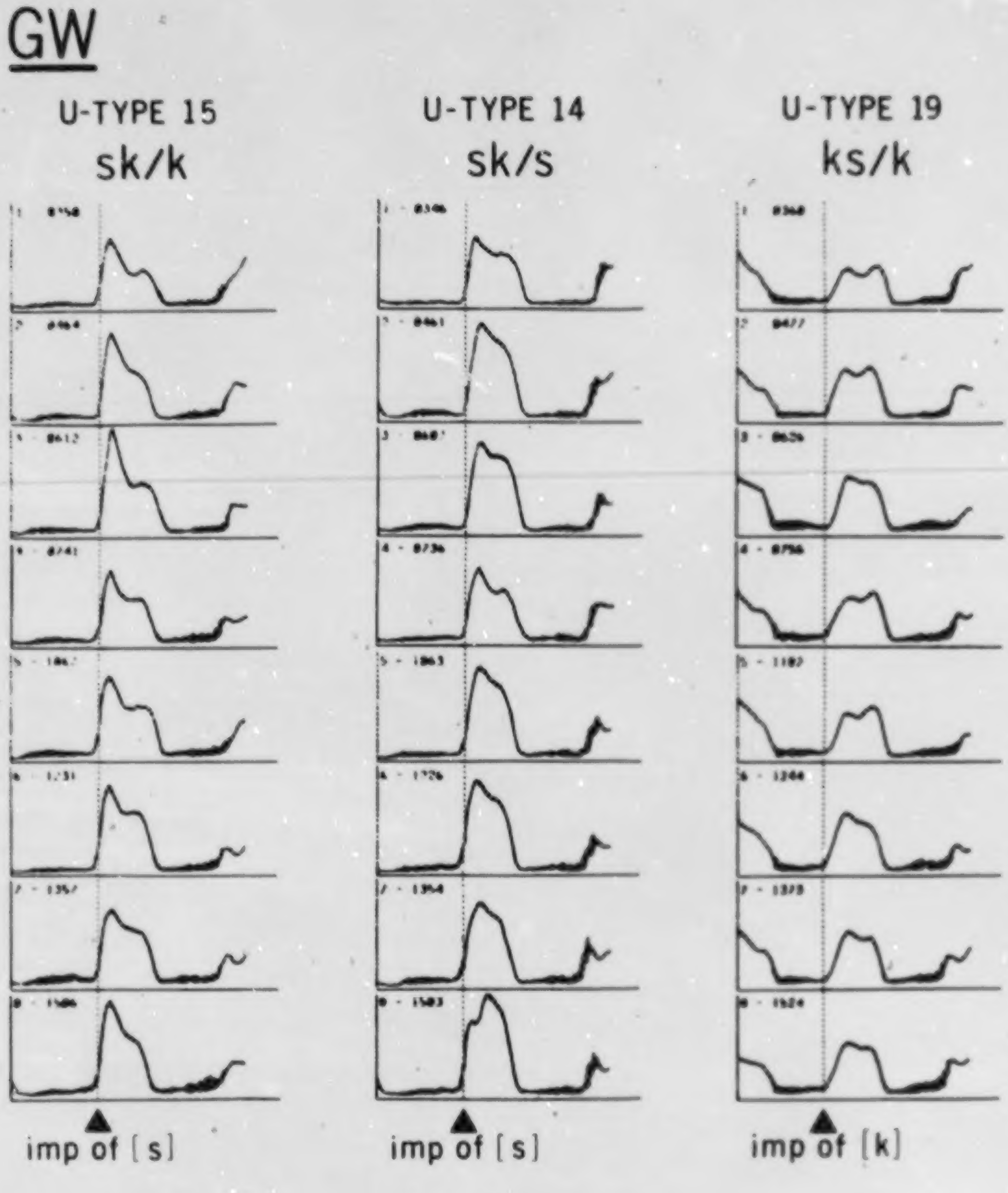
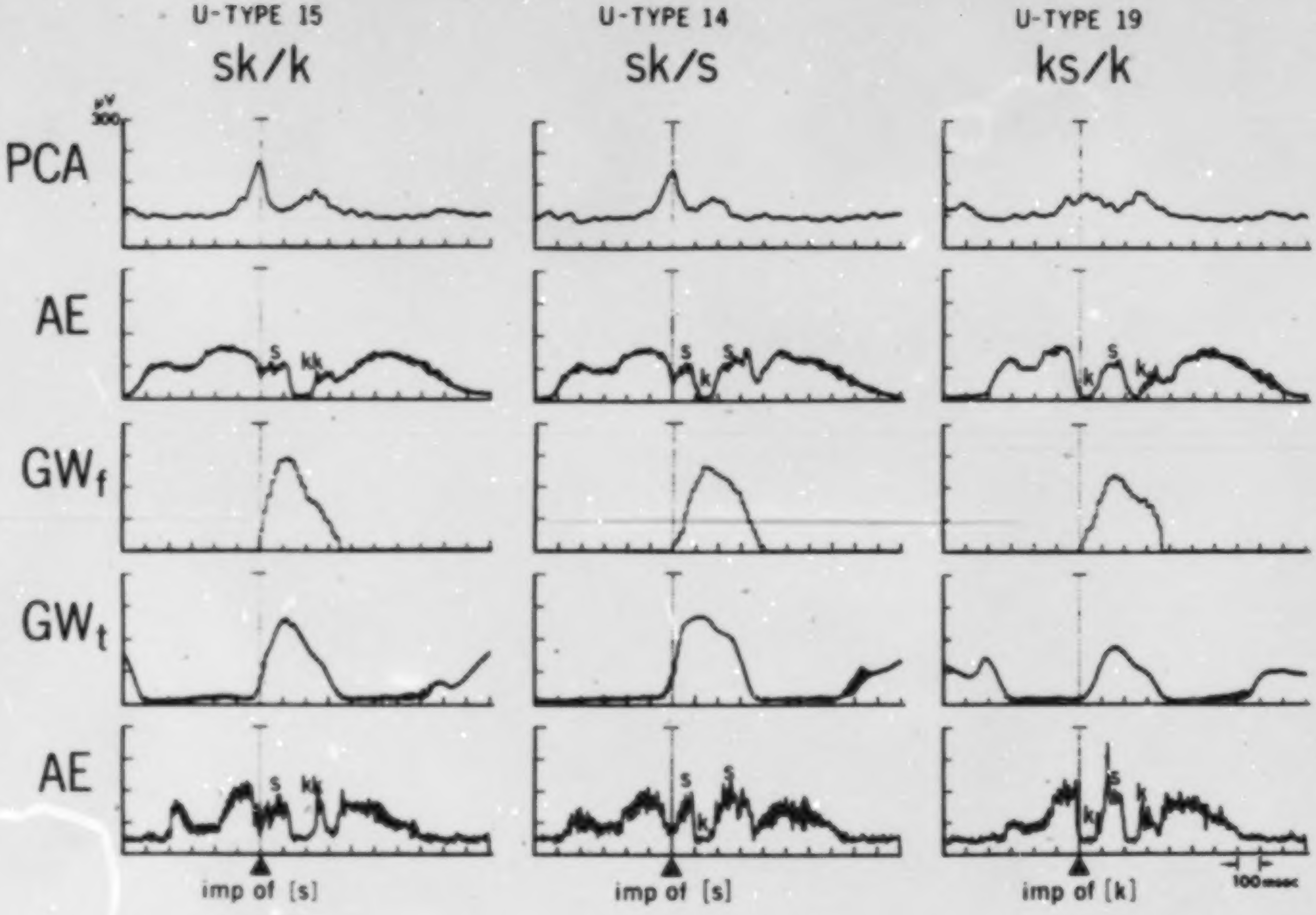


Figure 8: Averaged electromyograms of PCA, averaged audio envelopes, representative plots of glottal width using fiberoptics, corresponding glottograms and audio envelopes for same three utterance types as in Figure 7.



opening of the glottis. It is also evident that peak glottal opening is attained during either the fricative [s] segment or the aspirated [kh] segment, regardless of utterance type, which is consistent with the previous results for the various /sk/ combinations.

In addition, an inspection of these figures might lead one to assume that the two peak values in both EMG and movement curves for each utterance type are in a rather regular order across the speech material; in most cases, the first is bigger than the second, although this is not the case for averaged PCA curve for utterance type 19 in Figure 8, the glottogram for the token #8 of utterance type 14 in Figure 7, nor those for the token #1, 2, 4, and 5 of type 19 in Figure 7. Furthermore, closer inspection reveals that the first maximum value of glottal opening during the word final /ks/ production of utterance type 19 is smaller than those for the word final /sk/ sequences of types 15 and 16, despite the fact that peak timing in type 19 is usually later than those in types 15 and 16. The small peak opening for the word-final /ks/ sequence in type 19 is, therefore, most likely to be correlated with the slower velocity of the opening glottal movement.

Figures 9 and 10 contain three similar combinations, each composed of four phones. The variability in the number of peaks tends to increase with the number of the sequential voiceless phones; therefore we show here only the single token patterns of PCA activity in addition to those of glottography. These figures further indicate that each voiceless obstruent specified by aspiration or frication noise tends to be accompanied by a single separate opening gesture, while an unaspirated stop in a voiceless environment can be produced within the opening gesture attributed to the adjacent aspirated stop or fricative. Thus, both EMG and movement curves of the first eight productions for these three utterance types seem to be characterized by one, two, and three peaks, respectively, although identification is complex and uncertain in some cases. Observation of the velocity of the glottographic curves further reveals that the initial opening phase is slower for the voiceless sequence beginning with [k] in type 20 than those for the clusters beginning with [s] in types 16 and 23, resulting in a difference in the magnitude of the first peak between these two groups; the maximum opening for [ks] in type 20 is generally smaller than for [sk] in types 16 and 23. These findings are reasonably comparable to those described for Figures 7 and 8.

DISCUSSION

The technique of photo-electric glottography, as a tool for estimating glottal area variations both in vibration and articulation by registering the amount of light passing through the glottis, has been extensively applied to speech research since its introduction (Sonesson, 1960). Although it is intrinsically impossible to calibrate the system as a whole (Sawashima, 1974), some potential sources of error--such as light blockage by the tongue body or the epiglottis, displacement of the instruments inside the pharynx, and fogging of the fiberscope tip--can be minimized, or at least detected, by simultaneous fiberoptic monitoring of the glottal image through the optical view finder.

Furthermore, as is shown in Table 2, the correlation coefficients between the photo-electric amplitude and the value of the glottal width measured on

Figure 9: Glottographic patterns for first 8 productions of three utterance types containing various combinations of four voiceless phones.

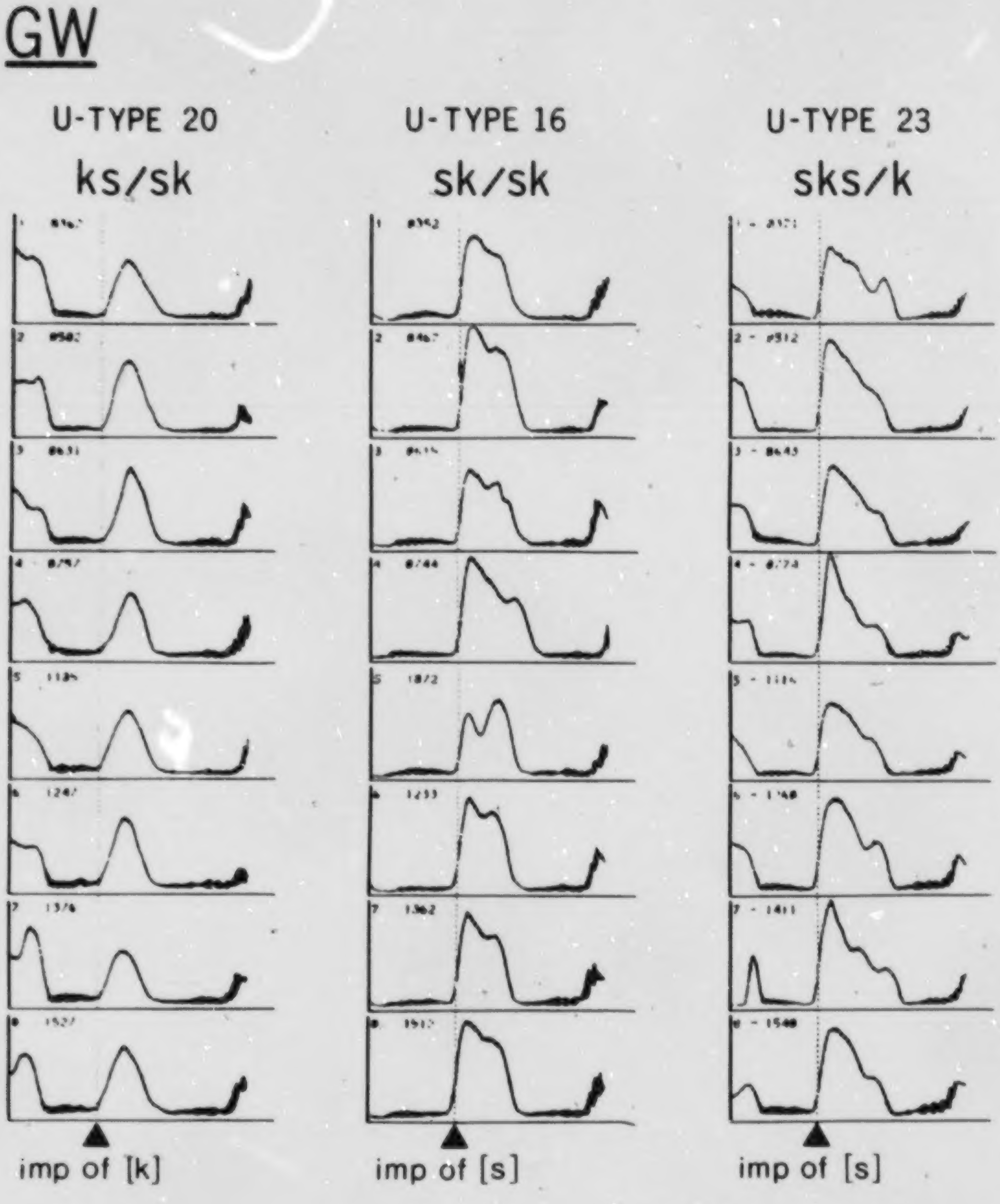


Figure 10: PCA activity patterns for first 8 productions of same three utterance types as in Figure 9.

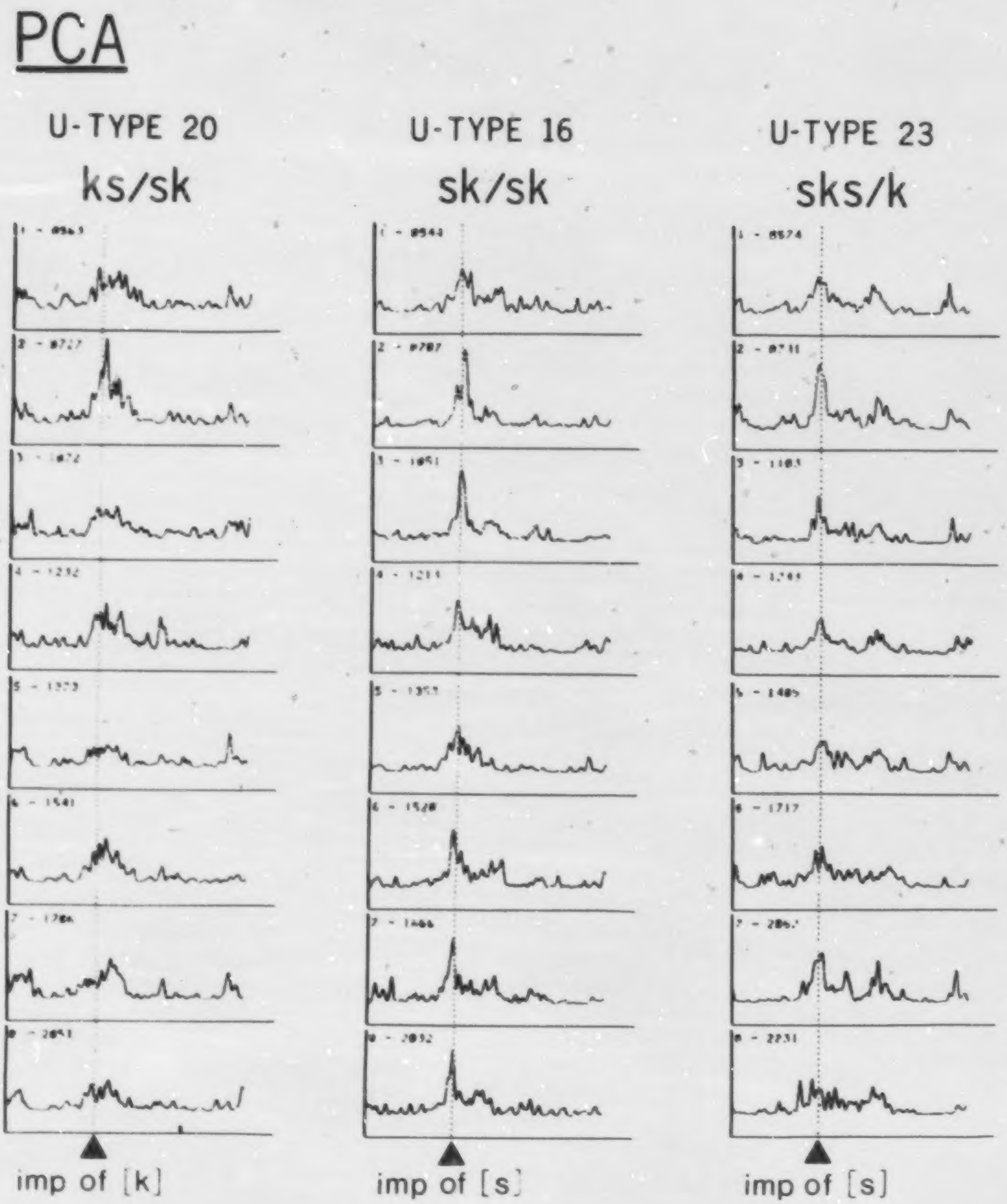


Table 2

Correlation coefficients between photo-electric signal and fiberoptic measurement of the glottal width during voiceless sequence production. (Two tokens of each utterance type.)

U-TYPE.	r_1=	(N_1)	r_2=	(N_2)	U-TYPE.	r_1=	(N_1)	r_2=	(N_2)	U-TYPE.	r_1=	(N_1)	r_2=	(N_2)
1.	------	(--)	------	(--)	9.	+0.975	(33)	+0.795	(33)	17.	+0.740	(40)	+0.922	(33)
2.	+0.876	(50)	+0.951	(50)	10.	+0.997	(53)	+0.922	(63)	18.	+0.817	(66)	+0.968	(73)
3.	+0.964	(33)	+0.927	(42)	11.	+0.935	(63)	+0.853	(77)	19.	+0.966	(73)	+0.948	(70)
4.	+0.878	(53)	+0.921	(56)	12.	+0.928	(66)	+0.825	(63)	20.	+0.928	(83)	+0.915	(70)
5.	+0.764	(20)	+0.765	(13)	13.	+0.906	(40)	+0.644	(47)	21.	+0.875	(63)	+0.862	(60)
6.	+0.946	(60)	+0.905	(63)	14.	+0.991	(73)	+0.728	(80)	22.	+0.803	(70)	+0.774	(83)
7.	+0.878	(57)	+0.976	(60)	15.	+0.958	(70)	+0.916	(83)	23.	+0.920	(83)	+0.690	(93)
8.	+0.914	(66)	+0.996	(70)	16.	+0.914	(83)	+0.935	(90)	24.	+0.766	(93)	+0.833	(80)

the projected image of simultaneously recorded fiberoptic films are highly positive during the opening and closing phases of the glottal movement, ranging from +0.644 to +0.997. However, even the fiberoptic method can not provide an accurate absolute value of the glottal area. In addition, the fiberoptic technique also has inherent problems; among others, the slow filming rate due to the limitation of the current light source power, the rather poor quality of the image, and the time consuming data reduction process. The authors conclude that simultaneous recording using both techniques is necessary to obtain reasonable estimations of glottal area variations during gross abduction and adduction, and that visual monitoring through the fiberoptic endoscopy permits reliable interpretation of large amounts of photo-electric recordings. [The other kind of problem, encountered in observing the cycle-to-cycle vibrations (Coleman & Wendahl, 1968; Harden, 1975), is not considered here.]

There are a few experimental reports on laryngeal articulation during clusters and/or geminates of voiceless obstruents, but the spoken samples in most are limited. Frøkjaer-Jensen et al. (1971) stated that the /s#p/ sequence in Danish subjects showed a two-peaked shape in slower pronunciation but only a single peak in normal speech. Lindqvist (1972) observed glottographic patterns during the Swedish word initial /sp/ sequence and found that there was no abduction gesture specific to the stop production. Fujimura and Sawashima (1971) described the glottalization of morpheme final [t] in the /t#t/ or /t#d/ sequence in American English in terms of false cord approximation, based on qualitative analysis of fiberoptic filming. Sawashima and Niimi (1974) reported that, using four Japanese subjects, the glottal opening gesture during voiceless segments, including geminates, showed a rather simple pattern with a single peak, together with some individual variations--particularly in the peak values. The materials did not contain voiceless clusters due to the fact that the phonology of Japanese does not allow voiceless "pure" clusters other than geminates. Löfqvist (1978) presented simultaneous recordings of glottograms and certain aerodynamic parameters during selected Swedish obstruent sequences and demonstrated that the peak glottal opening was found to occur during the fricative in clusters of stop + fricative and fricative + stop. Pétursson (1978) investigated the cluster production in Icelandic, showing the two peaks for [s + th], one for [st +].

The current results are generally in good agreement with those findings, despite the facts that the languages are different and that the number of subjects is usually limited. It was found in this experiment that the glottal opening gesture during a sequence of voiceless obstruents in American English is organized in a one, two, or more than two peaked mode, depending on the segmental properties of the phones used. The peaks were always reached during voiceless obstruents that were specified only by aspiration or frication noise. No particular opening was detected for an unaspirated stop. Rather, unaspirated stops seemed to be properly produced within the opening gesture attributed to the adjacent aspirated stop or fricative, during which segment glottal opening was maximized. These findings appear to be correlated with the aerodynamic requirement for obstruent production; a definitely separated glottis during these segments is indispensable for the egressive air flow that provides the source of aspiration or frication noise (Stevens, 1971), while this is not necessarily the case for the unaspirated stop.

An overall view of the entire EMG and movement data might be of interest. Figure 11 shows the relationship between the first maximum opening of the glottis and the first peak value of the PCA activity during each of the 23 different combinations of voiceless obstruents /s/ and /k/, using the averaged glottographic curves and the averaged EMG activity patterns, respectively. It should be mentioned here that the scattered characters "s" and "k" stand for voiceless sequences beginning with /s/ and /k/, respectively, such as "s" for /sk#k/ and "k" for /ks#k/. That is, the point "k" means that the X-Y coordinates correspond to the first peak value of the two parameters for the combination that was initiated by /k/ regardless of the following consonant(s), even if the peaks were actually reached during a following [s] segment if any. Nevertheless, it is clearly shown here that, in addition to the highly positive correlation as a whole, the peak values of both parameters seem to be categorized according to the manner of the initial voiceless obstruents; the first maximum opening as well as the peak abductor activity are generally larger for the sequences proceeding from [s] than those from [k]. Note that most of the "k" points, except for the three that correspond to the single word initial, word final and geminated /k/, are always reached during the following [s] segments. Therefore, it is conceivable that these peak values are more closely linked, not to the segments during which the peaks are reached, but to the initial segments from which those peaks start being reached, as far as the first opening phase during sequential voiceless obstruent production is concerned. Incidentally, the two points "s" embedded in the "k" group correspond to the word final /s/ in /s#V/ and /s#k/, while the lowest valued "k" is the word final /k/ in /k#V/.

Figure 12 presents the timing of the first maximum glottal opening during voiceless sequence production, using the averaged glottograms of 12 tokens each. In addition, two representative time courses of the original averaged glottograms are shown by the dashed lines during their initial opening movements up through the first peaks. The characters "s" and "k" are labelled according to the method used for Figure 11. From this figure, we may conclude that the difference in the first peak opening found in the previous figure is mainly related to the difference in the velocity of the glottal movement. That is, the clusters beginning with an [s] segment are accompanied by rapid initial opening, consequently attaining an early and larger maximum value of glottal aperture. On the other hand, in the clusters beginning with [k], the movement is gradual up to the first peak, even though the peak itself is reached during the following [s] segments, if any exist. It thus appears that a rapid opening of the glottis is necessary for the turbulence noise source during fricative segments; for stop production, however, such a rapid increase in glottal area seems unnecessary during initial stop closure to terminate vocal fold vibrations and prepare for following aspiration or frication noise, if required.

The current data might also be viewed from a more linguistic viewpoint, i.e., the phonetic effect of a word boundary. Let us assume for the moment that the word boundary in English is manifested at the laryngeal level by a closing movement at the preceding word ending, followed by an opening movement at the following word initiation. A similar hypothesis was once proposed by Fujimura (1972), although it referred solely to Korean stop production. Our speculation could explain, for example, why the glottis temporarily narrows in the vicinity of the juncture during /s#k/ production and reopens for the

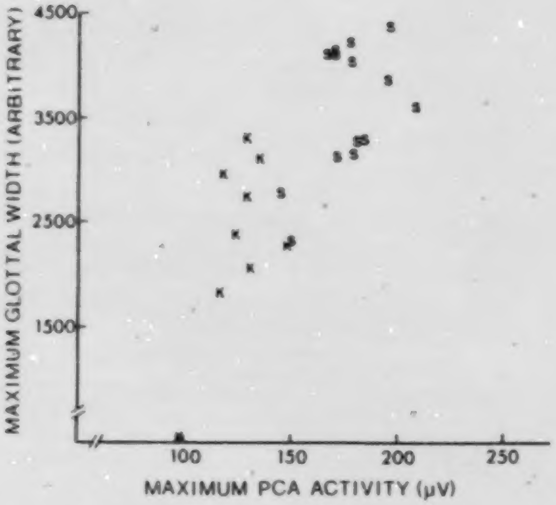


Figure 11: The relationship between the first maximum PCA activity and the first maximum glottal opening during voiceless sequence production, using the averaged curves of 12 tokens of each utterance type. "s" and "k" stand for the voiceless sequence begun with /s/ and /k/, respectively.

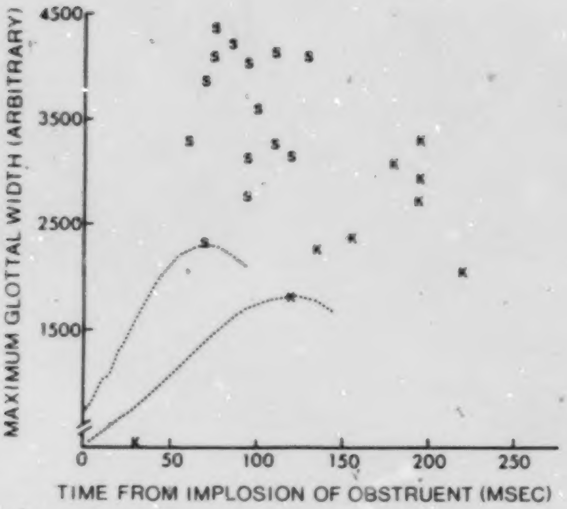


Figure 12: The timing of the first maximum glottal opening during voiceless sequence production, using the averaged curve of 12 glottograms of each utterance type. In addition, two representative time courses in the opening phase are included on the dashed lines. The time axis 0 msec corresponds to the implosion of the first obstruent, which served as line-up point for the averaging. "s" and "k" stand for voiceless sequence begun with /s/ and /k/, respectively.

following word-initial stop aspiration. Moreover, not only the glottalization of the word-final /k/ and the glottal attack of the word-initial vowel production might be accounted for along these lines, but also the difference in the peak glottal opening between the word-final /s/ and the word-inital /s/ production. Of course, so speculative an interpretation has several apparent defects; for instance, it does not explain the findings for the geminate cases, which were produced with a single opening gesture despite the word boundary. In addition, the peak value differences between the two opening gestures in the glottograms (shown in Figure 9) and electromyograms (shown in Figure 10) for /sk#sk/ production are rather contradictory, in that the peak values for the preceding word-final /sk#/ are usually larger than those for the following word initial /#sk/.

Another biomechanical interpretation is that there is an upper and lower limit on the velocity of glottal opening and closing gesture during speech. Here, the lower limit implies that static open positions of the glottis do not occur in running speech, and the glottal area is therefore continuously changing. The upper limit simply means that the velocity of glottal abduction and adduction movements can not exceed a certain value. Under this hypothesis, the general pattern of multiple glottal opening and closing gestures, found typically during a long cluster of voiceless obstruents, would be due to the lower limit, i.e., glottal articulation would be "cyclical" in nature. On the other hand, the mono-modal pattern observed in geminate production would be explained by the upper limit. If this assumption is eventually demonstrated as correct across subjects, speaking rate should substantially affect the number of peaks found in the glottal opening curve during sequential production of unvoiced sounds, as was put forth by Frøkjaer-Jensen et al. (1971).

In conclusion, the material presented in this paper has shed some light upon the nature of laryngeal control in the sequential production of voiceless obstruents. The observations will be used as guide-lines for further studies directed toward constructing a more comprehensive model of laryngeal articulation. In view of the fact that speaker-specific characteristics intersect with those that are language-specific, the authors are collecting more data on different speaking rates using several subjects and languages.

REFERENCES

Basmajian, J., & Stecko, G. A new bipolar indwelling electrode for electromyography. Journal of Applied Physiology, 1962, 17, 849.

Coleman, R., & Wendahl, R. On the validity of laryngeal photosenser monitoring. Journal of the Acoustical Society of America, 1968, 44, 1733-1735.

Dixit, R. P. Neuromuscular aspects of laryngeal control: With special reference to Hindi. Unpublished doctoral dissertation, University of Texas, Austin, 1975.

Fant, G., & Scully, C. (Eds). The larynx and language. Phonetica, 1977, 34, 245-324.

Frøkjaer-Jensen, B., Ludvigsen, C., & Rischel, J. A glottographic study of some Danish consonants. In L. L. Hammerich, R. Jakobson, & E. Zwirner (Eds.), Form & substance. Copenhagen: Akademisk Forlag, 1971, 123-140.

Fujimura, O., & Sawashima, M. Consonant sequence and laryngeal control.

Research Institute of Logopedics and Phoniatrics, University of Tokyo, Annual Bulletin, 1971, 5, 1-6.

Fujimura, O. Acoustics of speech. In J. H. Gilbert (Ed.), Speech and cortical function. New York: Academic Press, 1972, 107-165.

Harden, R. J. Comparison of glottal area changes as measured from ultra high speed photographs and photo electric glottographs. Journal of Speech and Hearing Research, 1975, 18, 728-738.

Hirano, M., & Ohala, J. Use of hooked-wire electrodes for electromyography of the intrinsic laryngeal muscles. Journal of Speech and Hearing Research, 1969, 12, 362-373.

Hirose, H. Posterior cricoarytenoid as a speech muscle. Annals of Otology, Rhinology and Laryngology, 1976, 88, 334-343.

Hirose, H., & Gay, T. The activity of the intrinsic laryngeal muscles in voicing control: An electromyographic study. Phonetica, 1972, 25, 140-164.

Hirose, H., & Gay, T. Laryngeal control in vocal attack: An electromyographic study. Folia Phoniatrica, 1973, 25, 203-213.

Hirose, H., Gay, T., Strome, M., & Sawashima, M. Electrode insertion technique for laryngeal electromyography. Journal of the Acoustical Society of America, 1971, 50, 1449-1450.

Iwata, R., & Hirose, H. Fiberoptic acoustic studies of Mandarin stops and affricates. Research Institute of Logopedics and Phoniatrics, University of Tokyo, Annual Bulletin, 1976, 10, 47-60.

Jakobson, R., Fant, C. G. M., & Halle, M. Preliminaries to speech analysis: The distinctive features and their correlates. Cambridge, Mass.: MIT Press, 1951.

Kagaya, R. A fiberoptic and acoustic study of the Korean stops, affricates and fricatives. Journal of Phonetics, 1974, 2, 161-180.

Kagaya, R., & Hirose, H. Fiberoptic, electromyographic and acoustic analysis of Hindi stop consonants. Research Institute of Logopedics and Phoniatrics, University of Tokyo, Annual Bulletin, 1975, 9, 27-46.

Kewley-Port, D. EMG signal processing for speech research. Haskins Laboratories Status Report on Speech Research, 1977, SR-50, 123-146.

Lindqvist, J. Laryngeal articulation studied on Swedish subjects. Quarterly Progress and Status Report, Speech Transmission Laboratory, Royal Institute of Technology, Stockholm, 1972, 2/3, 10-27.

Lisker, L., Abramson, A., Cooper, F. S., & Schvey, M. H. Transillumination of the larynx in running speech. Journal of the Acoustical Society of America, 1969, 45, 1544-1546.

Löfqvist, A. Closure duration and aspiration for Swedish stops. Working Papers, Phonetic Laboratory, Department of General Linguistics, Lund University, 1976, 13, 1-39.

Löfqvist, A. Laryngeal articulation and junctures in the production of Swedish obstruent sequences. In E. Garding, R. Bruce, & R. Bannert (Eds.), Nordic prosody. Lund: Department of Linguistics, 1978.

Pétursson, M. Joinure au niveau glottal. Phonetica, 1978, 35, 65-85.

Sawashima, M. Glottal adjustments for English obstruents. Haskins Laboratories Status Report on Speech Research, 1970, SR-21/22, 187-200.

Sawashima, M. Devoicing of vowels. Research Institute of Logopedics and Phoniatrics, University of Tokyo, Annual Bulletin, 1971, 5, 7-13.

Sawashima, M. Laryngeal research in experimental phonetics. In T. A. Seboek et al. (Eds.), Cuurent trends in linguistics, Vol. 12, Linguistics and adjacent arts and sciences. The Hague: Mouton, 1974, 2303-2348.

Sawashima, M. Current instrumentation and techniques for observing speech organs. Technocrat, 1976, 9-4, 19-26.

Sawashima, M., & Hirose, H. A new laryngoscopic technique by use of fiberoptics. Journal of the Acoustical Society of America, 1968, 43, 168-169.

Sawashima, M., & Niimi, S. Laryngeal conditions in articulations of Japanese voiceless consonants. Research Institute of Logopedics and Phoniatrics, University of Tokyo, Annual Bulletin, 1974, 8, 13-17.

Sonesson, B. On the anatomy and vibratory pattern of the human vocal folds. Acta Oto-Laryngologica, 1960, Suppl. 156.

Stevens, K. Airflow and turbulent noise for fricative and stop consonants: Static considerations. Journal of the Acoustical Society of America, 1971, 50, 1180-1192.

Yoshioka, H. Laryngeal adjustments during Japanese fricative and devoiced vowel production. Haskins Laboratories Status Report on Speech Research, 1979, SR-58, 147-160.

FOOTNOTE

[1] The deceptive shortness of the stop closure period, for example, in the audio envelope corresponding to each EMG curve should be attributed to the averaging method, while the continuous noise in the audio signals for the movement data is mainly due to the motor of the cine camera.

LEXICAL TONE AND SENTENCE PROSODY IN THAI*

Arthur S. Abramson+

Abstract. In speech production, the phonologically distinctive tones of a tone language are characterized primarily by fundamental frequency (F_0) contours and levels. The question arises as to whether sentence intonation, which is also described mainly as variations in the F_0 of the voice, is likely to weaken or even destroy the phonetic integrity of lexical tones. The ideal shapes of the five tones of Standard Thai (Siamese) are taken to be the ones that appear in citation forms of monosyllabic words. The shape of a tone embedded in an utterance can be perturbed somewhat through coarticulation with consonants and tones in the immediate context. Beyond that, F_0 measurements of running speech in Thai show a complicated interaction of lexical tones and sentence prosodies. In non-emotive speech, three terminal pitch junctures, found at major syntactic breakpoints, carry much of the sentence intonation; these junctures frequently occur with particles in which the lexical tones are then lost. This is not to be confused with tone changes occurring on frequently used function words. Elsewhere in the sentence, the full system of five tones seems to be preserved, although their ideal shapes undergo much distortion in running speech.

BACKGROUND

In a true tone language, one in which, in principle, every syllable in the morpheme stock bears a distinctive tonal phoneme, the tones are characterized primarily by fundamental-frequency levels and contours. Since we also describe intonation mainly in terms of the fundamental frequency of the voice, there appears to be a paradox involved in examining the relations between sentence prosody and word prosody in a tone language. As in other languages, so also in tone languages, is there the possibility of expressing attitudes or indicating certain aspects of syntactic structure by means of sentence intonation. The question arises as to whether the effects of sentence

---

*In Proceedings of the Ninth International Congress of Phonetic Sciences, 1979. (Copenhagen: University of Copenhagen), 380-387. Invited paper prepared as part of the author's participation in Symposium No. 7, "The Relation between Sentence Prosody and Word Prosody" in Copenhagen.

+Also University of Connecticut

Acknowledgment: This work was supported by NICHD Grant HD-01994 and BRS Grant RR-05596.

intonation are strong enough to weaken--or even destroy--the phonetic integrity of lexical tones.

The citation form of a monosyllabic word may be viewed as bearing the ideal manifestation of a tone. Of course, except for the occasional one-word sentence, such ideal forms do not often occur in running speech; yet children in the culture often learn new words this way, and so do adults in a foreign-language class. Once we have two or more tone-bearing syllables strung together, we expect perturbations through coarticulation. The final physical shaping of a tone is provided by the intonation of the utterance (Pike, 1948, 18-19).

THE TONES OF THAI

The ideal shapes of the tones of Standard Thai (Siamese) have been described elsewhere (Abramson, 1962; Erickson, 1974). It is useful to divide the five distinctive tones of the language into the "dynamic" class, comprising the rising and falling tones, and the "static" class, comprising the high, mid and low tones. The dynamic tones show rapid movements of F_0, while the static tones show rather slow movements which sometimes approximate F_0 levels. Of the three static tones, it is the mid tone that is most likely to appear occasionally as a level. The high tone is more likely to be seen as a rise high in the voice range in contrast with the low rise of the rising tone. The low tone is likely to appear as a low fall in contrast with the high fall of the falling tone.

Two types of phonetic context perturb the ideal shapes of the tones. Voiceless initial consonants induce a higher start of the F_0 contour, while voiced initial consonants induce a lower start (Gandour, 1974; Erickson, 1974). This kind of perturbation seems to have little effect on the phonetic integrity of the five tones, although it may serve as a supplementary cue to the voicing state of the initial consonant. It has been argued by historical linguists (Li, 1977), with some perceptual support from recent experiments on Thai (Abramson & Erickson, 1978), that through the phonemicization of these perturbations, the tones of Proto-Tai increased from three to the present-day sets of five or more in the modern languages of the family.

The phonetic context that causes greater deviations from the ideal tonal shapes is that of neighboring tones. In a series of tones spoken without pauses, tonal coarticulation occurs. Although physiological studies of Thai tones (Erickson, 1976) have yet to be extended to sequences, we can infer from acoustical evidence (Abramson, 1979) that this kind of coarticulation is manifested through the overlap of the effects of motor commands for the control of the laryngeal tensions and aerodynamic forces used.

Two sequential effects must be discriminated from tonal coarticulation. First, certain unstressed CV syllables with short /ɑ/ which have low or high tones in citation form are phonemically toneless, normally, in running speech. Another view is that the high and low tones on these syllables are neutralized, and the resulting pitch is assigned to the mid tone. This conclusion is handy for transcription, but the physical evidence suggests instability with

F_0 values dominated by the contours of the neighboring lexical tones. The other sequential effect to be excluded from consideration is tonal sandhi. The phonology of the language dictates that when certain kinds of morphemes are conjoined to form compound words, the lexical tone of one of the morphemes is replaced by another tone.

SENTENCE INTONATION AND TONES

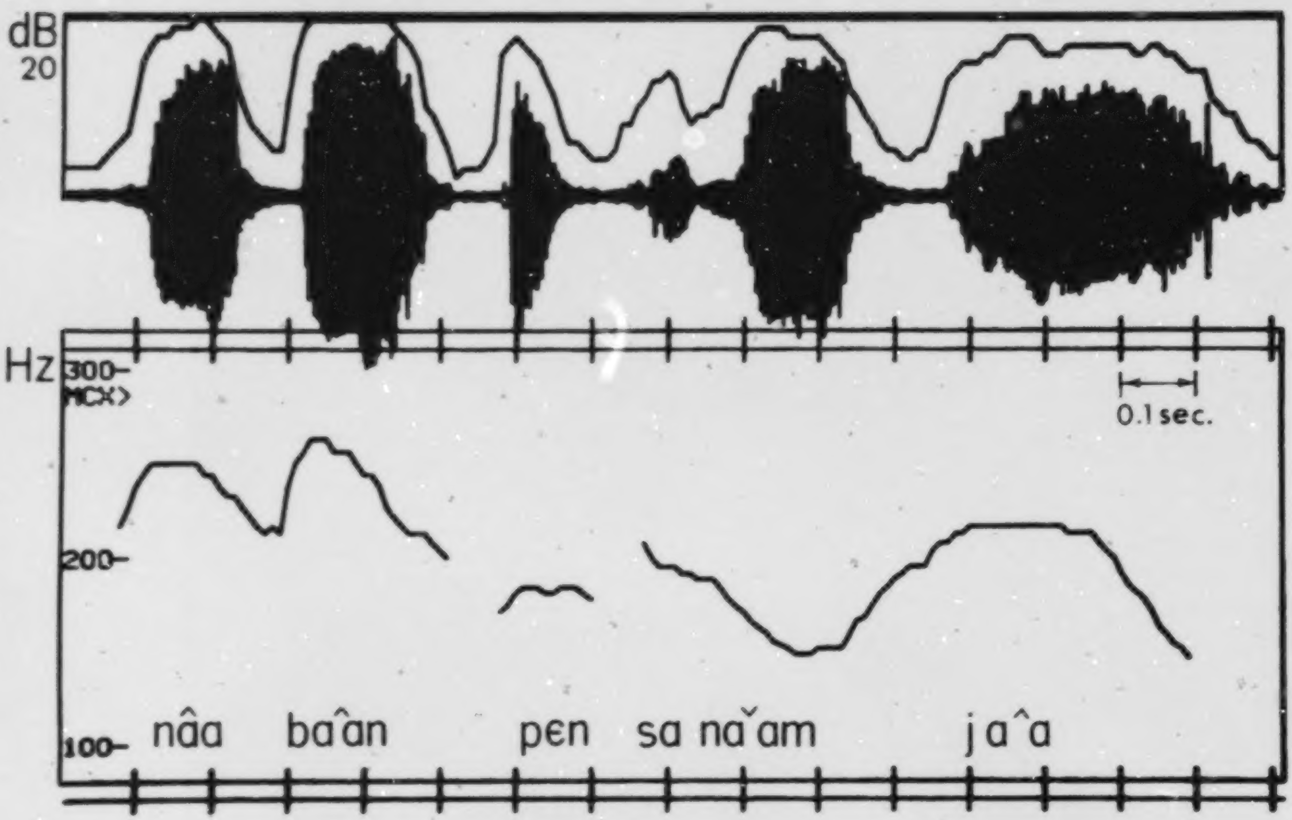
As one listens to spoken Thai, whether it be an animated conversation or a phlegmatic technical explanation, it becomes clear that, in addition to emotional states, such linguistic features as sentence accent and signs of major syntactic breakpoints can be expressed prosodically. The distinction between a statement and a question can also be expressed. In my present approach to the topic, I must lean mainly on my own extensive auditory but limited instrumental observations, as very few useful insights are found in the literature. It would be helpful if native Thai linguists or phoneticians gave more attention to the matter.

As a data base for such observations as I am ready to make, I have used two kinds of speech material. One is a conversation between two Thai adults of about one minute in length, recorded by J. Marvin Brown for a textbook published by the American University Alumni Association Language Center in Bangkok, Thailand. The other is a minute-long monologue recorded of a dean at a Bangkok university as he talked about a new academic program.

A computer-implemented analysis yielded displays of root-mean-square amplitude, wave forms, and F_0 contours. Cepstral analysis was used to extract the fundamental frequency. A sample set is shown in Figure 1 for the female speaker in the dialogue. Here, by the way, can be seen an example of tonal coarticulation. The phrase /nâa bâan/ 'in front of the house' bears two falling tones. The F_0 of the first one does not fall as far as the second; this presumably facilitates the resetting of the larynx for the sharp rise and fall of the second falling tone.

To handle the non-emotive aspects of sentence prosody in Thai, my examination of the present corpus of utterances, reinforced by the arguments of Rudaravanija (1965), leads me to posit three terminal junctures: rising pitch, sustained pitch, and falling pitch. These junctures function at clause ends and sentence ends. They may also function wherever the speaker pauses. The presence of a juncture affects the phonetic shape of the lexical tone on the last one or two syllables. The rising and falling junctures are likely to appear at the end of a breath group. In earlier work (Abramson, 1962) I also posited two pitch registers, high and normal, as units for Thai intonation. I now doubt the relevance of such registers for the non-emotive aspects of sentence prosody in the language. Indeed, to capture emotive prosodic variation, a somewhat more elaborate scheme might be needed. Although, as shown by Noss (1972) and Thongkum (1976), rhythmic factors play a role in Thai sentence prosody, they are excluded here because of the scope set by the organizers of the Congress.

Figure 1. From top to bottom: R.M.S. amplitude, wave form and fundamental frequency of a Thai woman's production of the sentence /nâa bâan pen sanâam jâa/ 'There is a lawn in front of the house'.



Henderson (1949) has argued that aside from the general melodic line of Thai intonation, the "sentence tone" as a whole is mainly determined by the speaker's choice of particles, most of them final particles. She describes seven such sentence tones. Without entering into the question of how many sentence tones there might be, I can at least say that these particles (which indicate, e.g., the sex of the speaker and something about the social relation between the speaker and the hearer) are prime carriers of the terminal junctures. Each particle as a lexical item has a tone of its own in citation form; this tone is usually predictable from the spelling. I doubt, however, that in running speech these "lexical" tones have any standing. The actual pitch imposed on a particle or, sometimes, a sequence of two particles, seems to be determined by the intonation of the whole sentence culminating in a terminal juncture. The resulting "tones" on these particles can sometimes be aligned with the lexical tones of Thai phonology; more often they are deviant. Some linguists, apparently in the grip of the view that every Thai syllable must bear a phonemic tone, feel constrained to write each particle with one of the five tones.

In both colloquial and formal discourse, many a sentence contains no particles, so the terminal junctures appear on the final word of the clause or sentence. Figure 1 shows such an effect. The falling tone on /jâa/ 'grass' at the end of the sentence is considerably lower both at its high point and low point than the two falling tones at the beginning. Even the rising tone just before it on /sanǎam/ 'field' does not rise to a point much higher than the immediately preceding mid tone on /pen/ 'be'. With such a short utterance it is hard to decide whether we have a final falling juncture on the compound word for 'lawn' or a falling intonation contour on the whole sentence.

Sentence accent is manifested by one or more of the following factors: (1) lengthening of the syllable, (2) a tonal contour that approaches the form of the ideal tone, and (3) an increase in amplitude. In the sentence in Figure 1 the final syllable appears to bear the sentence accent, using factors (1) and (2). In the phrase /nâa bâan/ at the beginning of the sentence, the second syllable is stressed, using factors (1) and (3); the amplitude trace is flattened at the top of the available 20-dB range, indicating saturation.

The points made so far have been descriptions of gross F_0 contours. A problem in intonation analysis is how to present quantitative data that go beyond overall "tunes." The prosodic constructs of the linguist often elude the measuring devices of the phonetician. With a simple-minded analysis for non-emotive prosody into three terminal junctures as a framework, I have made an initial tabulation of frequency movements for such clear examples of terminal juncture as I could find in the corpus. To provide for reasonable comparability of speakers, I treated frequency shifts at terminal junctures as percentages of the voice range. The maximum and minimum F_0 values for each of the three speakers are given in Table 1. Although the speech in both samples was calm, the narrower range for the monologue may not be due so much to the habits of that speaker as to the rather dispassionate and thoughtful nature of his discussion compared to the more animated dialogue.

Table 1

Voice Range in Hz

	Dialogue		Monologue
Speakers	A*	B**	U.W.**
Spread:	130-290	90-235	85-160
Range:	160	145	75

* Woman
** Man

The juncture of sustained pitch is generally found at syntactic breaks where the overall pitch of the voice neither rises nor falls before a brief pause; with or without a pause, the final syllable is prolonged. I have used this sustained pitch as a neutral reference from which to track the movements of the other two junctures. Examining both samples by ear and by eye, I accepted as valid tokens of the three junctures only those instances that were quite unambiguous. This cautious procedure yielded the small number of data in Table 2. The juncture of rising pitch signals surprise, doubt or a question. (Questions can also be marked by means of particles and other morphemes without terminal rising pitch.) The terminal fall appears at the ends of sentences and some major clauses. The "shift" for the sustained pitch is set at 0% as a neutral reference level, while the other two junctures are entered as departures from the neutral level. The data are averaged across three speakers. None of the tokens of these junctures happened to occur with the low lexical tone.

Table 2

Average Shift Through Voice Range for Terminal Pitch Junctures

Rising		Sustained		Falling	
N	%	N	%	N	%
6	30	14	0*	27	25

* Neutral reference point

Even away from the junctures, intonation has great effects on the realizations of the tonal phonemes. If the ideal forms of the tones have any psychological validity, then the forms in the sample of running speech have undergone severe distortion. A full account is beyond my reach here. At the same time, as I look at the contours and listen to the speech, I find preservation of the full system of five tones in running speech. That is, the usual linguistic scheme is not an artifact of the formal analysis of the linguist concentrating on citation forms only. Excluded from this generalization, however, must be all particles occurring at major syntactic breaks; they generally have their pitch determined by the sentence intonation without the involvement of lexical tones. Other frequently used function words, such as modals and pronouns, often undergo tonal replacement.

CONCLUSION

The phonemic tones and sentence prosodies of Thai interact in a rather complicated fashion. Three terminal pitch junctures, often occurring on particles, carry much of the intonation. Although the lexical tones are much influenced in their F_0 movements by sentence intonation, the contrasts between them are preserved except for certain small sets of morphemes. Sentence prosody allows for sentence accent. As in non-tonal languages, it is possible in Thai to use pitch junctures for differentiating between statements and at least some kinds of questions.

REFERENCES

Abramson, A. S. *The vowels and tones of standard Thai: Acoustical measurements and experiments*. Bloomington, Ind.: Indiana University Research Center in Anthropology, Folklore and Linguistics, 1962.

Abramson, A. S. The coarticulation of tones: An acoustic study of Thai. In T. L. Thongkum, P. Kullavanijaya, V. Panupong, & K. Tingsabadh (Eds.), *Studies in Tai and Mon-Khmer phonetics and phonology in honour of Eugénie J. A. Henderson.* Bangkok: Indigenous Languages of Thailand Research Project, 1979.

Abramson, A. S., & Erickson, D. M. Diachronic tone splits and voicing shifts in Thai: Some perceptual data. *Haskins Laboratories Status Report on Speech Research*, 1978, *SR-53*, Vol. 2, 85-96.

Erickson, D. Fundamental frequency contours of the tones of standard Thai. *Pasaa*, 1974, *4*, 1-25.

Erickson, D. M. *A physiological analysis of the tones of Thai*. Unpublished doctoral dissertation, University of Connecticut, 1976.

Gandour, J. Consonant types and tone in Siamese. *Journal of Phonetics*, 1974, *2*, 337-350.

Henderson, E. J. A. Prosodies in Siamese: A study in synthesis. *Asia Major*, 1949, *1*, 189-215.

Li, F-K. *A handbook of comparative Tai*. Honolulu: University Press of Hawaii, 1977.

Noss, R. B. Rhythm in Thai. In J. G. Harris & R. B. Noss (Eds.), *Tai phonetics and phonology*. Bangkok: Central Institute of English Language, 1972, 33-42.

Pike, K. L. *Tone languages*. Ann Arbor, Mich.: University of Michigan Press, 1948.

Rudaravanija, P. *An analysis of the elements in Thai that correspond to the* basic intonation patterns of English. Unpublished doctoral dissertation, Teachers College, Columbia University, 1965.

Thongkum, T. L. Rhythm in Thai from another view point. Pasaa, 1976, 6, 144-158.

INFLUENCE OF VOCALIC ENVIRONMENT ON PERCEPTION OF SILENCE IN SPEECH: AMPLITUDE EFFECTS

Bruno H. Repp

Abstract. This experiment investigated the influence of the relative amplitudes of the preceding and following vocalic portions on the perception of silence as a cue for the distinction between single and double stop consonants, both for nonidentical (/b/ vs. /b-g/) and identical (/b/ vs. /b-b/) places of articulation. The effects were generally small and in the direction of increased double-stop responses as amplitude increased. In the case of nonidentical places of articulation, only the preceding vocalic portion had a significant effect, whereas both vocalic portions had independent effects in the case of identical places of articulation. These results supplement those of Repp (1979) concerning effects of spectrum and duration of vocalic context; together, they place constraints on the form that a theory of silence perception in speech might take.

INTRODUCTION

In earlier studies (Repp, 1979) I investigated the influence of spectrum and duration of preceding and following vocalic portions on the perception of silence in speech. Their effect was measured in terms of the amount of silence needed to perceive (on half of the trials) a sequence of two stop consonants whose places of articulation were cued by formant transitions into and out of the silent interval. In one condition, loosely referred to as the *single-cluster* condition, the two sets of transitions conveyed different places of articulation (/b-g/); in the other condition, the *single-geminate* condition, both were appropriate to the same place (/b-b/). Roughly 70 msec of silence are needed to hear both stops in /b-g/ (only /g/ is heard at shorter silences), and about 200 msec are needed to hear both stops in /b-b/ (only one /b/ is heard at shorter silences). For general introductions and for discussions of these effects, the reader is referred to Dorman, Raphael, and Liberman (1979), Repp (1978), and, of course, Repp (1979), whose studies the present experiment supplements.

The purpose of my earlier experiments (Repp, 1979), as well as of the present study, was not so much to decide between alternative hypotheses concerning the perception of silence in speech--although some tentative

---

Acknowledgment: This research was supported by NICHD Grant HD01994 and BRS Grant RR05596. Thanks are due to Patti Price who ran the experiment and tabulated the data.

[HASKINS LABORATORIES: Status Report on Speech Research SR-59/60 (1979)]

conclusions could be drawn--as to provide data that would form the basis for a future comparison with psychophysical observations on the perception of silence in nonspeech context. When we know more about the psychoacoustics of silence perception--and our knowledge is quite incomplete in that regard--such a comparison will certainly place a strong constraint on the form that an appropriate theory of the role of silence in speech perception might take.

In addition to spectrum and duration--the factors investigated earlier (Repp, 1979)--amplitude is an important auditory parameter. By investigating its influence on silence perception in speech, the present study extends and complements the earlier experiments. There was an additional motivation of the present study: In the earlier experiment on spectral effects, certain alterations in the amplitudes of the synthetic stimuli were made that may have confounded the results. The present experiment was expected to provide an estimate of the magnitude of such possible confounding, so that the earlier findings might be re-evaluated.

Method

Subjects. Twelve subjects participated. They included ten paid student volunteers with little experience in listening to synthetic speech, a research assistant with some listening experience, and the author, a seasoned subject. The results of all subjects were combined, as no qualitative differences were apparent.

Stimuli and procedure. Stimuli and procedure were identical with those of Repp (1979, Exp. II), except that variations in amplitude replaced variations in duration. Only the most important information will be given here. The stimuli were synthetic and consisted of a VC portion (/ib/, 190 msec long) followed by a variable period of silence and a CV portion (/ga/ or /ba/, 290 msec long). In the single-cluster condition (/ib-ga/), the silent interval varied from 15 to 115 msec in 10-msec steps. In the single-geminate condition (/ib-ba/), it varied from 115 to 315 msec in 20-msec steps. The relative amplitudes of the VC and CV portions were varied orthogonally in three 6-dB steps. The amplitude changes in each portion were +6, 0, and -6 dB relative to the baseline stimuli (Repp, 1979); the changes were implemented in the digitized wave forms before the test sequences were recorded. Each subject heard each of the 99 stimuli in each condition (9 amplitude combinations, 11 silence durations) 8 times in randomized order and identified the stop consonant(s) heard. Single-cluster and single-geminate conditions were presented as separate blocks in counterbalanced order.

Results

Figure 1 shows percentages of double-stop (cluster or geminate, depending on the condition) responses as a function of silence duration. Separate response functions are shown for the nine VC-CV amplitude combinations in the two conditions. The 50-percent cross-overs of these functions--about 70 msec for the single-cluster distinction and somewhat below 200 msec for the single-geminate distinction--are in good agreement with earlier data. The effect of VC amplitude can be seen within each panel, whereas the effect of CV amplitude extends vertically across panels.

169

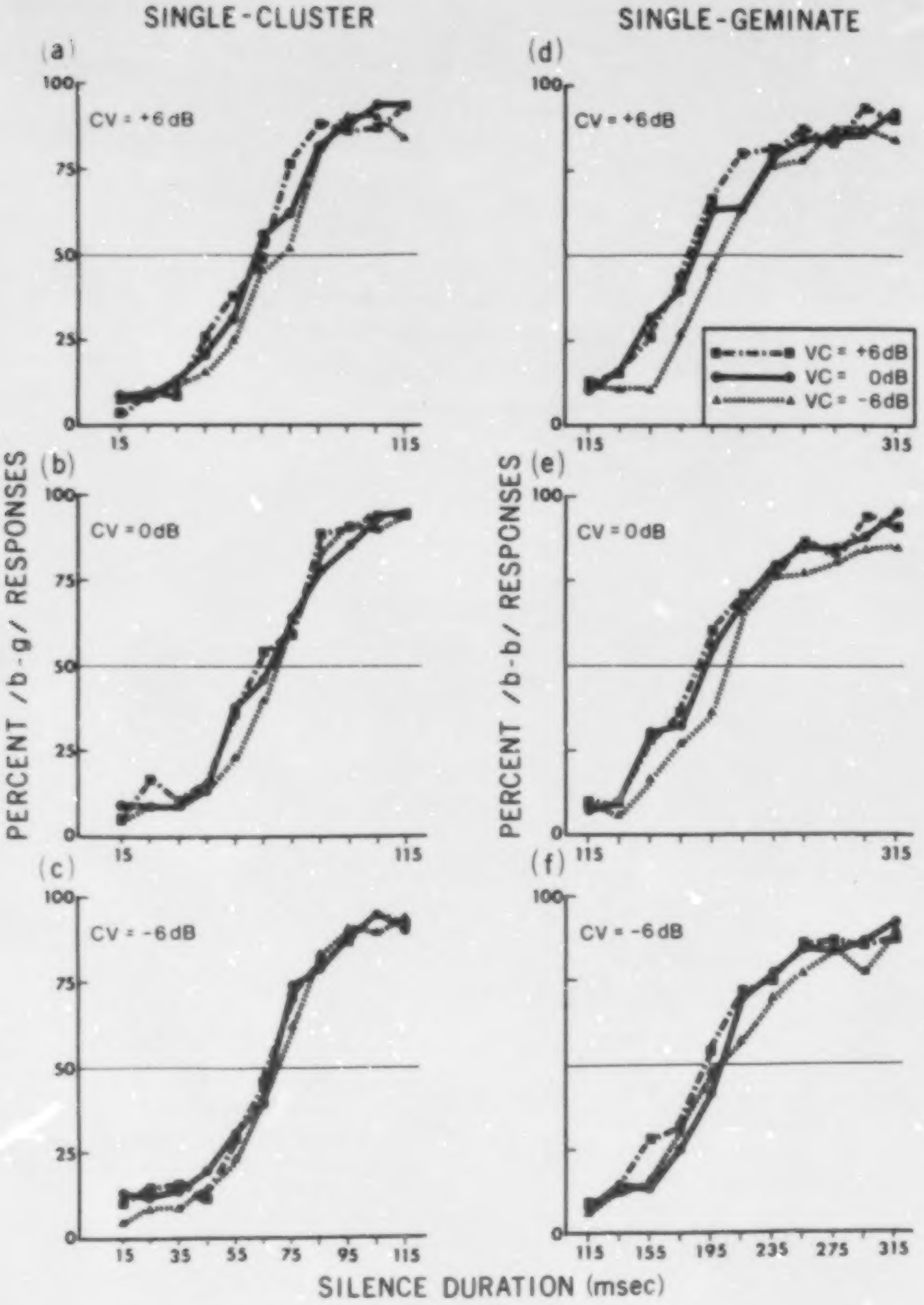


Figure 1. Percentage of double-stop responses as a function of silence duration for nine VC-CV amplitude combinations in two experimental conditions.

It is evident that amplitude effects were rather small, especially in the single-cluster condition. Nevertheless, the effect of VC amplitude did reach significance in that condition, $F(2,22) = 4.3$, $p < .05$. The overall percentage of cluster responses increased slightly with VC amplitude. There was no effect of CV amplitude here, and no interaction between VC and CV amplitude effects.

In the single-geminate condition, the effect of VC amplitude was somewhat larger and more consistent across different silence durations, as can be seen in Figure 1. It, too, was in the direction of more double-stop (here: geminate) responses as VC amplitude increased. The effect was highly significant, $F(2,22) = 10.2$, $p < .001$. In addition, there was a significant effect of CV amplitude, $F(2,22) = 7.2$, $p < .01$, again in the same direction. The VC and CV amplitude effects appeared to be independent, as there was no significant interaction. Note that this implies an increase in geminate responses with overall stimulus amplitude; thus, it was not the relative amplitudes of the VC and CV portions but their absolute levels that mattered.

Since some of these effects are difficult to see in Figure 1, the results are summarized in more concise form in Figure 2. Instead of category boundaries--which are difficult to estimate accurately from response functions with indeterminate asymptotes, such as those in Figure 1--Figure 2 simply plots the percentages of double-stop responses, averaged across all silence durations, as a function of the two amplitude parameters. Apart from the effects just discussed (now more clearly visible), it is evident that the effect of VC amplitude was nonmonotonic: a 6-dB attenuation had a larger effect than a 6-dB amplification.

DISCUSSION

These amplitude effects are small compared to those of changes in VC (and CV) spectrum and duration, especially in the single-cluster condition. Therefore, the pre-experimental amplitude changes perpetrated in the stimuli of the spectrum experiment (Experiment I of Repp, 1979) probably had little influence on the results. Moreover, the only substantial changes (9-10 dB attenuation) had been made in /ba/ and /ga/, both CV stimuli. Since the present experiment indicates little or no effect of CV amplitude over a 12-dB range, the earlier results are vindicated.

Consider now the theoretical implications of the present results. In my earlier paper (Repp, 1979), I discussed three hypotheses about the perception of silence in speech. According to the first, the backward masking hypothesis (which really applies to the single-cluster condition only), cluster responses should have increased with increases in the amplitude of the "target" (the VC portion), as indeed they did; however, one might also have expected them to decrease with increases in the amplitude of the "mask" (the CV portion), which they did not. Thus, the evidence is equivocal with respect to the backward masking hypothesis.

According to the second hypothesis, the articulatory hypothesis, the perceptual results should mirror what happens in speech production. Of course, simple amplitude changes of the sort used here hardly ever occur in natural speech production; therefore, the articulatory hypothesis cannot be

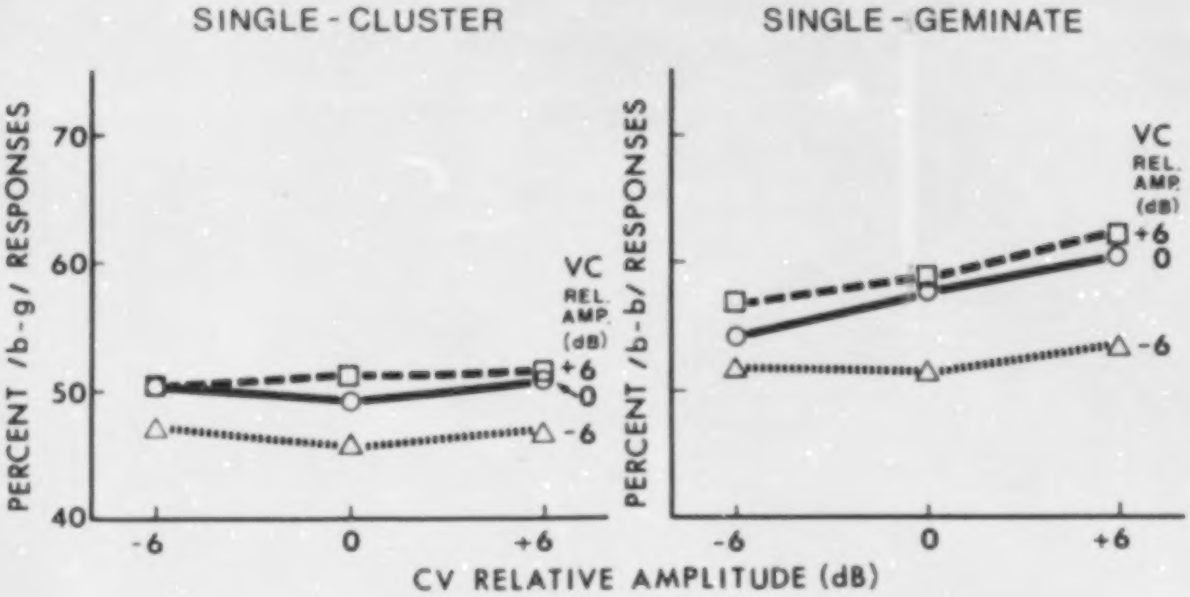


Figure 2. Percentage of double-stop responses, averaged over all silence durations, as a function of VC and CV amplitudes in two conditions.

easily evaluated here. Nevertheless, we note that amplitude changes in production are primarily associated with changes in stress, and the present experimental stimuli probably could have been judged by listeners to vary systematically in their stress pattern. However, perceived stress depends, especially when only a single parameter varies, on the relation between two signal portions, and the present results showed that this relation did not influence the subjects' responses. For this reason, the articulatory hypothesis seems not to be supported by the present data.

The third hypothesis, the differentiation hypothesis, claims that the listener's task is the perceptual separation of the auditory-phonetic events preceding and following the silence. One factor that facilitates separation is an increase in the effective (subjective or physical) duration of the silent interval. Increases in the amplitudes of the VC and CV portions presumably increased their distinctiveness as "markers" of this interval. Although it is not clear why this should increase the subjective duration of the silence (and there seem to be no relevant data from psychophysical studies), it might reduce uncertainty about the boundaries of the silence. There is evidence from the auditory literature that increases in the amplitude of brief, burst-like markers increase the discriminability of silent intervals (Abel, 1972; Carbotte & Kristofferson, 1973; Divenyi & Danner, 1977; Divenyi & Sachs, 1978). Thus, in the present study, the perceptual salience of the silent interval may have increased with increased signal amplitude, leading to better perceptual separation of the VC and CV portions, and of their associated phonetic messages. At the same time, the clarity of the signal components themselves may have improved. Thus, of the three hypotheses considered, the differentiation hypothesis seems to be most compatible with the present findings.

These conclusions are necessarily tentative. Psychophysical studies are planned to compare auditory and phonetic perception of silence more directly by matching auditory stimuli to the present VC-CV stimuli in duration, amplitude, and--as closely as possible--spectral characteristics. These studies should reveal whether the perception of silence in speech can be accounted for by auditory principles, or whether specifically phonetic processes must be postulated.

REFERENCES

Abel, S. M. Duration discrimination of noise and tone bursts. Journal of the Acoustical Society of America, 1972, 51, 1219-1223.

Carbotte, R. M., & Kristofferson, A. B. On energy-dependent cues in duration discrimination. Perception & Psychophysics, 1973, 14, 501-505.

Divenyi, P. L., & Danner, W. F. Discrimination of time intervals marked by brief acoustic pulses of various intensities and spectra. Perception & Psychophysics, 1977, 21, 125-142.

Divenyi, P. L., & Sachs, R. M. Discrimination of time intervals bounded by tone bursts. Perception & Psychophysics, 1978, 24, 429-436.

Dorman, M. F., Raphael, L. J., & Liberman, A. M. Some experiments on the sound of silence in phonetic perception. Journal of the Acoustical Society of America, 1979, 65, 1518-1532.

Repp, B. H. Perceptual integration and differentiation of spectral cues for

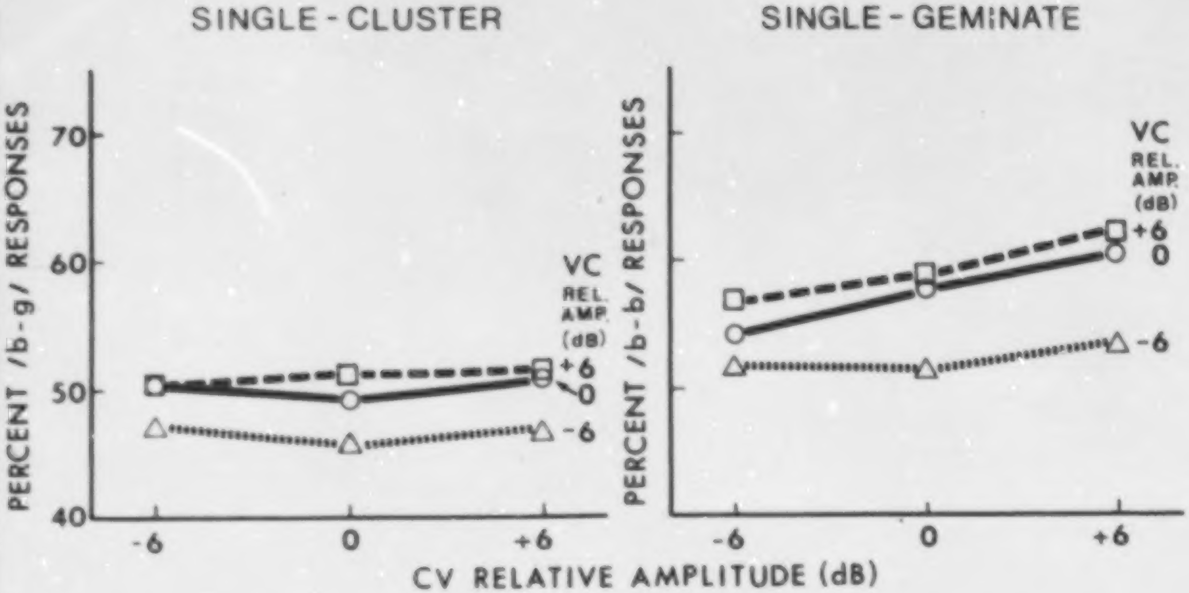


Figure 2. Percentage of double-stop responses, averaged over all silence durations, as a function of VC and CV amplitudes in two conditions.

easily evaluated here. Nevertheless, we note that amplitude changes in production are primarily associated with changes in stress, and the present experimental stimuli probably could have been judged by listeners to vary systematically in their stress pattern. However, perceived stress depends, especially when only a single parameter varies, on the relation between two signal portions, and the present results showed that this relation did not influence the subjects' responses. For this reason, the articulatory hypothesis seems not to be supported by the present data.

The third hypothesis, the differentiation hypothesis, claims that the listener's task is the perceptual separation of the auditory-phonetic events preceding and following the silence. One factor that facilitates separation is an increase in the effective (subjective or physical) duration of the silent interval. Increases in the amplitudes of the VC and CV portions presumably increased their distinctiveness as "markers" of this interval. Although it is not clear why this should increase the subjective duration of the silence (and there seem to be no relevant data from psychophysical studies), it might reduce uncertainty about the boundaries of the silence. There is evidence from the auditory literature that increases in the amplitude of brief, burst-like markers increase the discriminability of silent intervals (Abel, 1972; Carbotte & Kristofferson, 1973; Divenyi & Danner, 1977; Divenyi & Sachs, 1978). Thus, in the present study, the perceptual salience of the silent interval may have increased with increased signal amplitude, leading to better perceptual separation of the VC and CV portions, and of their associated phonetic messages. At the same time, the clarity of the signal components themselves may have improved. Thus, of the three hypotheses considered, the differentiation hypothesis seems to be most compatible with the present findings.

These conclusions are necessarily tentative. Psychophysical studies are planned to compare auditory and phonetic perception of silence more directly by matching auditory stimuli to the present VC-CV stimuli in duration, amplitude, and--as closely as possible--spectral characteristics. These studies should reveal whether the perception of silence in speech can be accounted for by auditory principles, or whether specifically phonetic processes must be postulated.

REFERENCES

Abel, S. M. Duration discrimination of noise and tone bursts. Journal of the Acoustical Society of America, 1972, 51, 1219-1223.

Carbotte, R. M., & Kristofferson, A. B. On energy-dependent cues in duration discrimination. Perception & Psychophysics, 1973, 14, 501-505.

Divenyi, P. L., & Danner, W. F. Discrimination of time intervals marked by brief acoustic pulses of various intensities and spectra. Perception & Psychophysics, 1977, 21, 125-142.

Divenyi, P. L., & Sachs, R. M. Discrimination of time intervals bounded by tone bursts. Perception & Psychophysics, 1978, 24, 429-436.

Dorman, M. F., Raphael, L. J., & Liberman, A. M. Some experiments on the sound of silence in phonetic perception. Journal of the Acoustical Society of America, 1979, 65, 1518-1532.

Repp, B. H. Perceptual integration and differentiation of spectral cues for intervocalic stop consonants. *Perception & Psychophysics*, 1978, *24*, 471-485.

Repp, B. H. Influence of vocalic environment on perception of silence in speech. *Haskins Laboratories Status Report on Speech Research*, 1979, *SR-57*, 267-290.

STIMULUS DOMINANCE AND EAR DOMINANCE IN FUSED DICHOTIC SPEECH AND NONSPEECH STIMULI: A REPLICATION

Bruno H. Repp

Abstract. The patterns of stimulus dominance and ear dominance effects were compared between two types of fused dichotic stimuli: two-formant CV syllables ranging from /bV/ to /dV/ to /gV/ (V = /a/ or /i/) and brief, isolated, steady-state resonances ("timbres") corresponding to the second-formant onset frequencies of the CV stimuli. Results for the two types of stimuli were similar; there was no evidence that the speechlike quality of the CV syllables influenced stimulus dominance or ear dominance effects, which thus seemed to be governed by auditory stimulus properties and individual differences in their perception. This result confirms earlier data of Repp (1978c).

INTRODUCTION

The present study replicates and extends an earlier experiment (Repp, 1978c), and the reader is referred to that report for a general introduction. Briefly, the purpose of the earlier study was to investigate whether the relative speechlikeness of a set of stimuli influences the pattern of stimulus dominance and ear dominance effects obtained in fused dichotic presentation. The answer was negative--stimulus dominance seemed to be governed by auditory stimulus properties (second-formant onset frequency), and the direction of ear dominance depended primarily on the individual listener, not on stimulus type. These findings provided evidence against the hypotheses that dichotic stimulus dominance reflects the relative "category goodness" of speech stimuli (Repp, 1976, 1977a, 1978a, 1978b), and that, within the range of stimuli considered, the right ear would become more dominant as the stimuli become more speechlike.

There were five types of synthetic stimuli in the earlier study, all derived from two-formant syllables ranging from /bæ/ to /dæ/ to /gæ/. Due to the small number of subjects and the difficulty of the task, the results for three of the stimulus types ("bleats", "transitions", and "chirps") were preliminary at best. However, more stable data were available for the two extreme (i.e., respectively most and least speechlike) stimulus sets--full CV syllables and "timbres" (40-msec steady-state, second-formant resonances at the frequencies that, as onset values, distinguished the CV syllables). It was the comparison between these, acoustically most dissimilar, sets that revealed the greatest similarity in response patterns.

Acknowledgment: This research was supported by NICHD Grant HD01994 and BRS Grant RR05596.

[HASKINS LABORATORIES: Status Report on Speech Research SR-59/60 (1979)]

CV syllables and timbres had in common the distinctive property of second-formant (F_2) onset frequency. The similarity of results for the two types of stimuli suggested that both stimulus dominance and ear dominance effects were governed by this auditory stimulus property (as well as by individual differences in its perception). To test the generality of this finding, the present study attempted to replicate it with two new sets of CV syllables and timbres. The new stimuli represented different ranges of F_2 onset frequencies, and, in the CV syllables, different vocalic contexts (/-a/, /-i/) than in the earlier stimuli (/-æ/). Shifts up or down the frequency scale were not expected to affect responses to steady-state timbres; this part of the study was almost certainly expected to replicate the earlier results. In CV syllables, on the other hand, the F_2 onset frequency changes rapidly toward a steady state characteristic of the vowel; depending on this vowel, a given onset frequency may initiate a rising, flat, or falling formant transition. If all that mattered in dichotic competition were F_2 onset frequency, then, of course, the nature of the transition and of the following vowel would be irrelevant. However, Pompino, Rilhac-Sutter, Simon, and Sommer (1977)--in a study that almost duplicates the CV condition of the present experiment but only recently came to my attention--report that the steepness of F_2 transitions is the decisive factor in dichotic competition. If so, quite different patterns of stimulus dominance relationships would be expected for the two sets of CV syllables described below. Moreover, it is well known that the perceived place of articulation of syllable-initial stop consonants is not determined by formant onset frequency alone, and to the extent that phonetic categorization influences dichotic stimulus dominance (as postulated by the "category goodness hypothesis" of Repp, 1976, 1977a, 1978a, 1978b), the two sets of CV syllables to be described should yield different response patterns. On the other hand, identical response patterns for all four stimulus sets (two CV series and two timbre series) would provide strong support for the sole importance of F_2 onset frequency in dichotic competition.

The predictions just discussed concern the relative dominance of one stimulus over another in dichotic competition. As to ear dominance, the questions were: Does ear dominance shift toward the right side (left hemisphere) as the stimuli become more speechlike? And are ear dominance effects for CV syllables and timbres related to each other? Repp's (1978c) earlier data suggested a negative answer to the first question and a positive answer to the second one, but because of the small number of subjects, a replication seemed desirable.

Method

Subjects. The eight subjects included six paid student volunteers, an undergraduate research assistant, and the author. All subjects had listened to synthetic speech before, but only the author had had extensive experience. A full replication of the author's data was available; these two sets of data were averaged before they were combined with those of the other subjects.

Stimuli. There were four sets of stimuli: /Ca/ syllables, /Ci/ syllables, and two corresponding sets of timbres. All stimuli were generated on the Haskins Laboratories parallel resonance synthesizer. Each set contained seven stimuli distinguished by F_2 onset frequency. The CV syllables were wholly periodic two-formant patterns with initial stepwise-linear formant transitions that led to the perception of /b,d,g/ preceding either /a/ or /i/. The F_1 transition was constant and 30 msec in duration; it went from 181 Hz to 743 Hz in /Ca/ syllables and from 181 Hz to 286 Hz in /Ci/ syllables. The F_2 transitions were 40 msec in duration and ended at 1075 Hz in /Ca/ stimuli and at 2307 Hz in /Ci/ stimuli. The variable F_2 onset frequencies are shown in Table 1. The portion following the formant transitions was perfectly steady-state, and total stimulus duration was 250 msec. To increase their speechlike quality, the steady-state portions of the CV syllables were given a linearly falling F_0 contour (from 114 Hz to 90 Hz), whereas the initial transitional portions had a constant F_0 of 114 Hz.

Table 1

F_2 onset frequencies (in Hz)

Stimulus	/Ca/ syllables Low timbres	/Ci/ syllables High timbres
1	921	1541
2	1075	1695
3	1232	1845
4	1386	1996
5	1541	2156
6	1695	2307
7	1845	2462

The two sets of timbres were matched to the two sets of CV stimuli. Timbres consisted of 40-msec steady-state F_2 segments at the frequencies listed in Table 1, with a constant F_0 of 114 Hz. Since the F_2 onset frequencies were lower in /Ca/ syllables than in /Ci/ syllables, the corresponding sets of timbres will be referred to as "low" and "high," respectively. Note that stimuli 5-7 of the low set were identical with stimuli 1-3 of the high set.

All stimuli were digitized at 10 kHz using the Haskins Laboratories pulse code modulation system. There were four experimental tapes, one for each of the four stimulus sets. Each tape contained first a binaural AXB discrimination test in which the five two-step pairings (1-3, 2-4, 3-5, 4-6, 5-7) of the stimuli within a set occurred three times in each of the four AXB configurations (AAB, ABB, BAA, BBA), yielding a total of 60 triads. This sequence was followed by a dichotic AXB test. In each dichotic AXB triad, X consisted of A and B presented simultaneously to the two ears, which resulted in strong fusion; A and B were presented binaurally. All fifteen pairings of stimuli two or more steps apart in a given series appeared in each of the four possible AXB configurations (A[AB]B, A[BA]B, B[AB]A, B[BA]A), yielding a total of 60 triads. Three different randomizations of these 60 triads were recorded. Finally, for each of the two CV syllable sets, there was a standard identification test containing twenty replications of each of the seven stimuli. The interstimulus intervals in this test, and between AXB triads, were 3 sec; those within AXB triads were 500 msec.

Procedure. Each subject listened to each of the four tapes in four separate sessions. The order was counterbalanced across subjects. The order of tests within conditions was as described above. The main purpose of the binaural AXB discrimination test was to familiarize the subjects with the stimuli. The task was to write down "A" when X equalled A, and "B" when X equalled B. In the dichotic AXB test, the task was the same, except that "was more similar to" replaced "equalled" in the instructions. The three dichotic sequences were repeated after a pause during which the channels-to-ears assignment was electronically reversed. Thus, each subject gave a total of 24 responses to each dichotic stimulus combination (disregarding ear assignment). In the final identification test (CV syllables only), the subjects identified the initial consonant in each stimulus as B, D, or G. Other details of procedure were the same as in Repp (1978c), as indeed were most of those described above.

Results and Discussion

Binaural identification and discrimination. The labeling and two-step discrimination results for the stimuli in the two speechlike series, averaged over subjects, are shown in Figure 1. It can be seen that the endpoint stimuli of the CV continua were rather consistently labeled B and G, respectively. However, no stimulus on either continuum was consistently labeled D, although D responses reached a maximum for stimuli 4 and 5. Presumably, alveolar stops need a third-formant transition or a burst (both absent in the present stimuli) to sound convincing. Two-step discrimination performance ranged from 62 to 90 percent. The precise pattern of discrimination results need not concern us here, but note that the discriminability of the individual stimuli set an upper limit to performance in the dichotic AXB test. However, that test contained not only two-step stimulus combinations--which may have presented some difficulty for the listeners--but also combinations of stimuli three to six steps apart; these, of course, were much easier to discriminate.

No labeling data were collected for the timbres, since they do not fall into natural categories. AXB discrimination performance for timbres was virtually perfect; only one subject made any errors at all.

Dichotic stimulus dominance. The average results of the dichotic AXB tests are plotted in Figure 2 as "percent 'i-ness' judgments ($i<j$)", i.e., the percentage of trials on which a fused dichotic stimulus was judged to be more similar to the component with the lower F_2 onset frequency. Individual stimulus pairs are identified by stimulus numbers in the graphs (i) and on the abscissa (j). Let us focus first on the results for timbres, shown on the left. First, it is evident that nearly all data points fall above the 50-percent (equilibrium) line. This indicates a strong tendency for low-frequency timbres to dominate high-frequency timbres, which replicates a similar trend found by Repp (1978c). As in the earlier study, this trend was shown by most, but not all, subjects. One subject, in particular, showed exactly the opposite (viz., high-frequency dominance), and to such an extent that her timbre data could not be used. Moreover, a second subject showed such strong low-frequency dominance for the set of low timbres that his data provided no information and were likewise excluded from the averages. Thus,

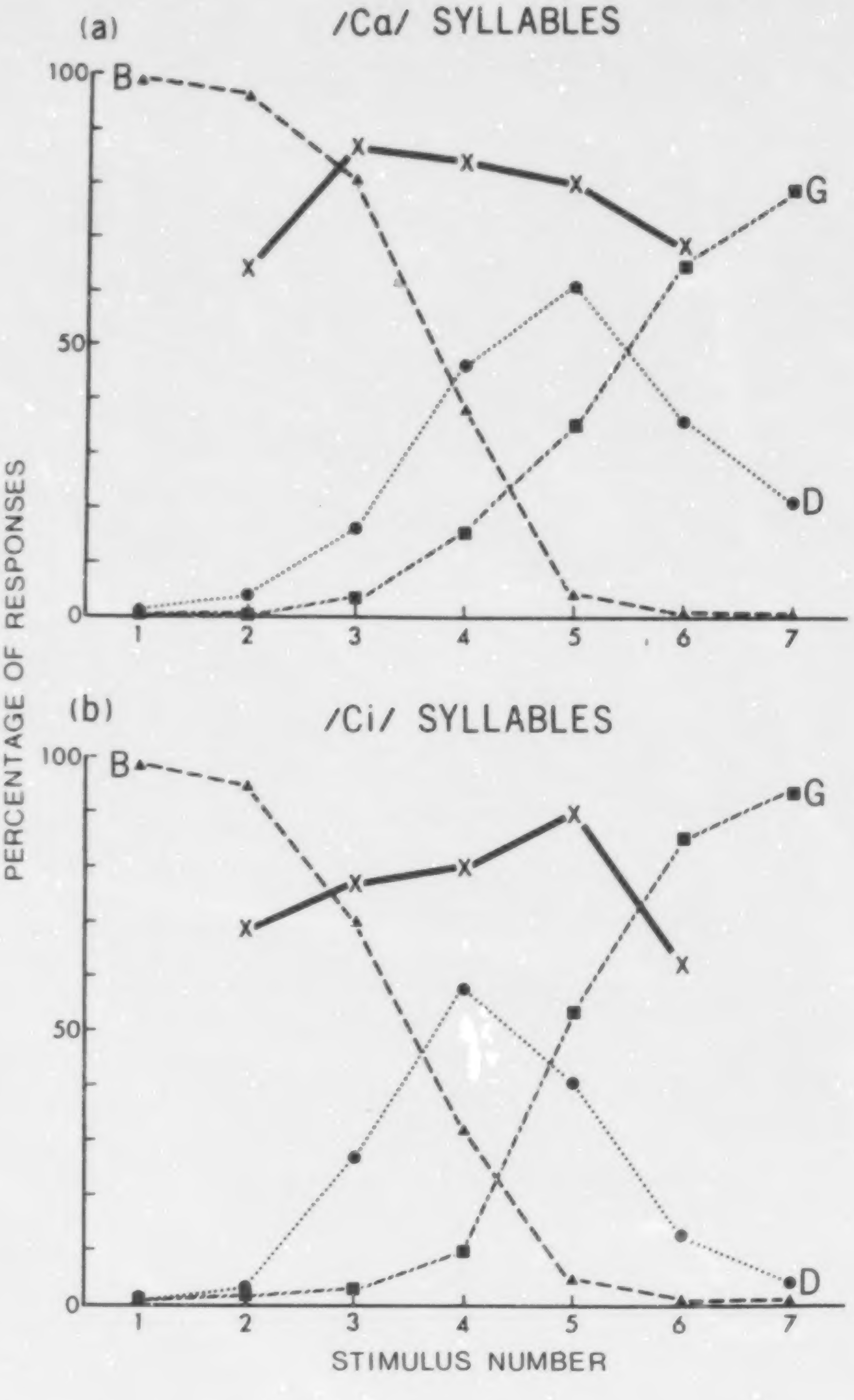


Figure 1. Labeling and 2-step discrimination functions for /Ca/ and /Ci/ stimuli.

the low timbre data in Figure 2 are based on only six subjects, whereas the high timbre data derive from seven subjects.

A second feature of the data to note is that stimulus pair 5-7 in the low set and stimulus pair 1-3 in the high set were physically identical but showed quite different results. In general, the fact that approximately equal degrees of overall low-frequency dominance were obtained in both sets of timbres (and in the timbre stimuli of Repp, 1978c) demonstrates that the degree of frequency dominance for a given stimulus pair was highly range-dependent. Obviously, listeners adapted to the frequency range of a given stimulus set and adjusted their criteria accordingly.

Third, we note that the main determinant of stimulus dominance within a given stimulus set was the frequency of the lower timbre in a pair: the lower its frequency, the more dominant it became. This is indicated by the vertical separation of the functions in Figure 2, which connect all points with a constant lower timbre. On the other hand, the frequency of the higher timbre in a pair had a much smaller effect, as indicated by the shallow slopes of the functions in Figure 2. Whatever effect there was, it too was in the direction of increased dominance as frequency changed from high to low.

Because of their peculiar structure, these data are difficult to analyze statistically. However, the consistency of the average response pattern is expressed in the (within-group) correlation across the two sets of timbres ($r = +0.84$, $p < .001$), and in their respective (between-group) correlations with the timbre data of Repp (1978c), obtained with stimuli in a frequency region precisely intermediate between the present two ranges ($r = +0.83$ and $+0.88$, both $p < .001$).

Turning now to the results for CV syllables, shown in the right half of Figure 2, we find a rather similar pattern. Overall low-frequency dominance was less pronounced, and so was the effect of the frequency of the stimulus with the lower F_2 onset, but both trends were clearly present, especially in /Ci/ syllables. The frequency of the higher-onset stimulus seemed to have no effect at all here (or, in /Ca/ stimuli, perhaps in a direction opposite to that observed in timbres). In general, the results for /Ca/ syllables showed no strong effects of any sort, indicating difficulties in discriminating and judging these stimuli. One factor that may have contributed to these difficulties was the tendency of simultaneous /ba/ and /ga/ to yield /da/ (cf. Cutting, 1976; Repp, 1976); this tendency was much less pronounced in /Ci/ stimuli, according to the author's observations as a subject. Of course, this phenomenon introduced random responses, since listeners did not (and in fact, had been encouraged not to) consistently judge /da/ to be more similar to either /ba/ or /ga/.

The important point about the CV data is that they provide little indication of a response pattern specific to their speechlike quality; in this respect, they replicate the data of Repp (1978c). From the category goodness hypothesis, one should have expected stimulus 7 (a good exemplar of /gi/ and a reasonable one of /ga/) to be a much stronger competitor than stimuli 3-6. There was a small trend in that direction--too small to be taken seriously. Nor do the data confirm the observation of Pompino et al. (1977) that the steepness of transitions determines their relative dominance. The reason for this discrepancy is not quite clear. However, the present results do support,

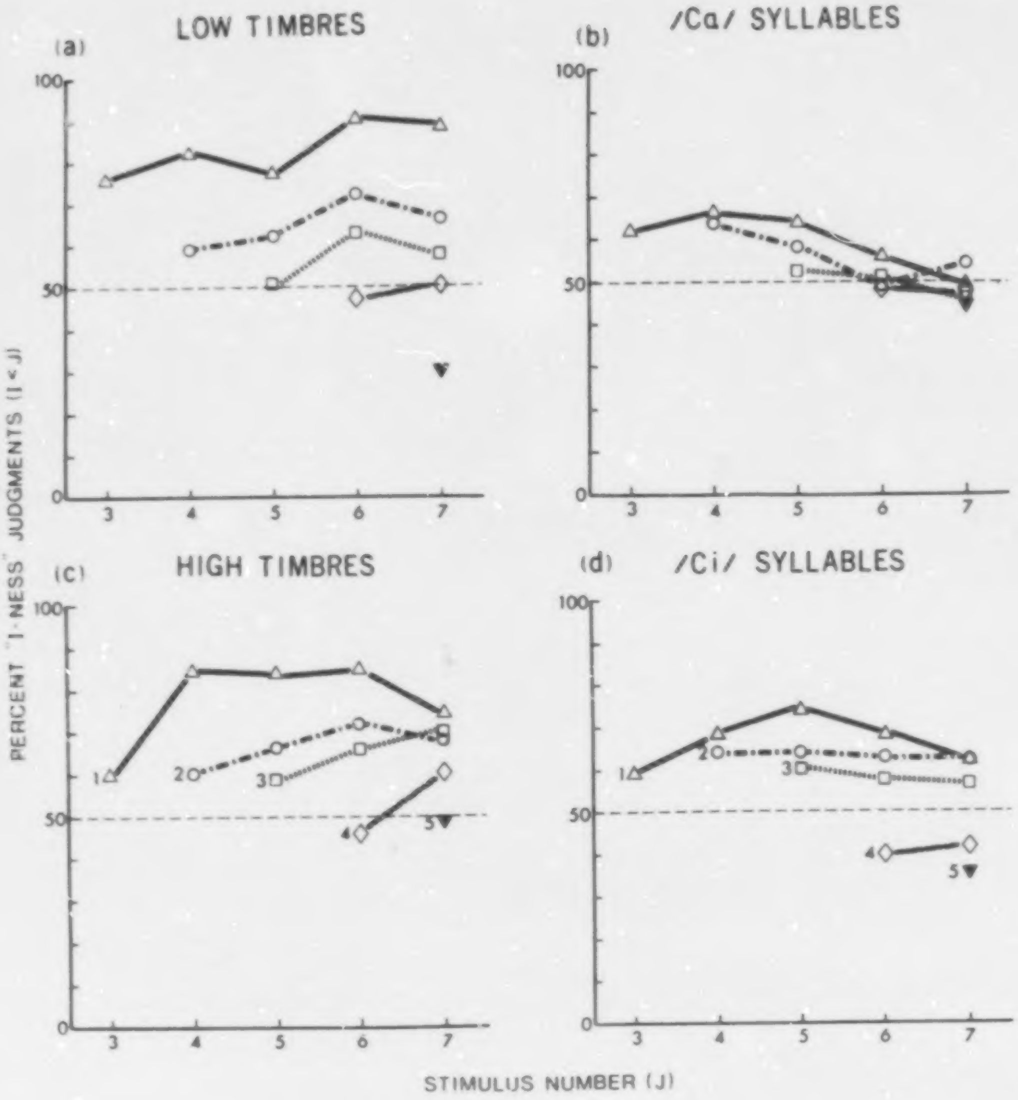


Figure 2. Stimulus dominance patterns for four types of stimuli.

in a rather indirect way, the suggestion (e.g., Stevens & Blumstein, 1978) that listeners are most sensitive to the onset spectrum of CV syllables.

The relative consistency of the CV data is indicated by the correlation (across the 15 data points) between the /Ca/ and /Ci/ results (r = +0.73, p < .001), as well as by at least one of their respective correlations with the earlier /C[illegible]/ data (r = +0.36, n.s., and r = +0.78, p < .001). Moreover, the response patterns for CV syllables and timbres were significantly similar in the present study (correlations between +0.48 and +0.82), as they had also been in Repp (1978c). Thus, it may be concluded that the stimulus dominance patterns for all stimulus series were essentially the same, all surface differences probably being due to variations in task difficulty.

Ear dominance effects. Individual ear dominance effects for the four stimulus sets are displayed in Table 2. The index shown is e' (Repp, 1977b), which ranges from -1 for perfect left-ear dominance to +1 for perfect right-ear dominance. As can be seen, there was a considerable number of significant effects, both in favor of the left ear and in favor of the right ear. What is not evident in the data is a consistent change in the direction of ear dominance as a function of stimulus type. Moreover, the ear dominance effects for all sets of stimuli seemed to be related (correlation coefficients between +0.51 and +0.78). Thus, the relative speechlikeness of the stimuli appeared to have no systematic influence on ear dominance. This once more confirms Repp (1978c).

Table 2

Ear dominance indices (e')

Subject	Low timbres	High timbres	/Ca/ syllables	/Ci/ syllables
1[a] (I)	+0.02	+0.81***	+0.34***	+0.26**
(II)	+0.20**	+0.56***	+0.16	+0.58***
2	+0.12**	+0.29**	-0.04	-0.41***
3	+0.21	-0.33***	+0.01	+0.06
4	+0.01	-0.14	-0.08	+0.11
5	-0.38**	-0.14	-0.11	-0.13*
6	-0.46***	-0.71***	-0.14*	-0.08
7	-----[b]	----[b]	+0.14*	-0.05
8	-0.41***	-0.77***	-0.27**	-0.55***

***p < .001
**p < .01
*p < .05

[a] The author; data from two sessions.
[b] No estimate because of extreme stimulus dominance effects.

The author, who was the only subject to show consistent right-ear dominance in all four sets of stimuli, also happened to be the only strongly right-handed subject. By a curious coincidence, all other subjects (drawn from a limited population of summer students) were either left-handed or had left-handed relatives. Perhaps, then, the hypothesis that ear dominance would be shifted to the right in CV syllables, relative to timbres, was not given a proper test, as some of the subjects may not have been left-hemisphere-dominant for speech. The subjects most likely to fall in that category were subjects 6 and 8, who were left-handers with left-handed relatives (Hardyck & Petrinovich, 1977). These subjects did show the largest left-ear dominance effects in the group; however, contrary to expectations, they were also the ones who showed most clearly a reduction of left-ear dominance with CV syllables--the opposite of what one should have expected if these subjects were right-hemisphere-dominant for speech. Of course, a reduction in absolute ear dominance was to be expected for CV syllables because of the greater difficulty of that condition. But subjects were not even consistent in that respect. Thus, the data continue to offer no consistent evidence of systematic variation in ear dominance among stimulus conditions. They are perhaps suggestive of a relation between ear dominance (for both timbres and CV syllables) and hemispheric dominance for speech, but this issue clearly requires further investigation, as does the possible relation between ear dominance for timbres and for pure tones contrasting in pitch (Efron & Yund, 1974).

Conclusion

The present study confirms that of Repp (1978c) in all aspects. It provides no convincing evidence that the relative speechlikeness of fused complex sounds contrasting in harmonic spectrum influences either stimulus dominance or ear dominance effects. Stimulus dominance seems to be a function of F_2 onset frequency and of the frequency range employed. Ear dominance depends primarily on the individual listener and on certain task variables, not on stimulus type; its relation to cerebral dominance for speech and/or pitch perception remains to be established.

REFERENCES

Cutting, J. E. Auditory and linguistic processes in speech perception. Psychological Review, 1976, 83, 114-140.

Efron, R., & Yund, E. W. Dichotic competition of simultaneous tone bursts of different frequency--I. Dissociation of pitch from lateralization and loudness. Neuropsychologia, 1974, 12, 249-256.

Hardyck, C., & Petrinovich, L. F. Left-handedness. Psychological Bulletin, 1977, 84, 385-404.

Pompino, B., Rilhac-Sutter, M., Simon, A., & Sommer, R. Auditorische Faktoren der Gewichtung bei psychoakustischer Fusion. Forschungsberichte (Institut fuer Phonetik und sprachliche Kommunikation, Universitaet Muenchen), 1977, 8, 97-120.

Repp, B. H. Identification of dichotic fusions. Journal of the Acoustical Society of America, 1976, 60, 456-459.

Repp, B. H. Dichotic competition of speech sounds: The role of acoustic stimulus structure. Journal of Experimental Psychology: Human Perception and Performance, 1977, 3, 37-50. (a)

Repp, B. H. Measuring laterality effects in dichotic listening. Journal of the Acoustical Society of America, 1977, 62, 720-737. (b)

Repp, B. H. Stimulus dominance in fused dichotic syllables. *Haskins Laboratories Status Report on Speech Research*, 1978, *SR-55/56*, 133-148. (a)

Repp, B. H. Categorical perception of fused dichotic syllables. *Haskins Laboratories Status Report on Speech Research*, 1978, *SR-55/56*, 149-162. (b)

Repp, B. H. Stimulus dominance and ear dominance in fused dichotic speech and nonspeech stimuli. *Haskins Laboratories Status Report on Speech Research*, 1978, *SR-55/56*, 163-180. (c)

Stevens, K. N., & Blumstein, S. E. Invariant cues for place of articulation in stop consonants. *Journal of the Acoustical Society of America*, 1978, *64*, 1358-1368.

USE OF FEEDBACK IN ESTABLISHED AND DEVELOPING SPEECH*

Gloria J. Borden+

INTRODUCTION

During spontaneous speech, we create utterances in our heads as we produce them out loud. As we formulate the next linguistic chunk to be spoken, we hold it momentarily, if only parts of it, as a perceptual image. We not only know what it is, in general, that we plan to say, but we have a rather abstract idea of how it will sound. Unconsciously, we also have an image of how it will feel in terms of touch, pressure sensations, and movements and positions of speech organs. We know these things because we know our own vocal tract possibilities and our own voicing capabilities. We have heard and felt ourselves talk for years. In almost every speaking situation we are able to feel and hear ourselves speak. We know so well how we will sound and feel when producing speech that we can continue to produce perfectly intelligible speech in artificial situations in which we are prevented from hearing ourselves, as under auditory masking, or are prevented from feeling surface sensations, as under oral anesthesia. It is likely that in these instances we continue to receive information on our performance from our muscles and from feedback mechanisms contained within the central nervous system.

Less skilled speakers must depend upon auditory and tactile feedback more than speakers who have well-established speech production systems. Young children developing speech, and speakers of any age attempting to learn a new language, must use all available feedback channels in their efforts to match the sound patterns of the new language with the sensations produced by their own imitations. Children with congenital hearing losses are evidently at a serious disadvantage in developing natural sounding speech patterns. In contrast, those who have learned to speak before the onset of deafness evidence good speech with only slight deterioration of intelligibility gradually taking place after a period of time (Fry, 1966). The use one is able to make of feedback from one's own speech seems to vary, too, with age. Children first acquiring language, whether they are learning one or two languages, are particularly adept at matching their own speech to the models provided. They use feedback to emulate the segmental, intonational, and rhythmic characteristics of the languages to which they are exposed. Too often, however, older people learning a second language fail in their efforts to match the segmental and suprasegmental aspects of the language as spoken, even though they may have mastered the grammar and vocabulary. The auditory, tactile, and muscle-

---

*To be published in N. Lass (Ed.), Speech and language: Advances in basic research and practice (Vol. III). New York: Academic Press.

+Also at Temple University, Philadelphia.

Acknowledgment: This paper was partially supported by NIH grant NS 13617 to Haskins Laboratories.

moving images they have stored for their first language seem to supercede any new images, so adults speak the second language with an accent. The sound patterns of the first language persist in the second.

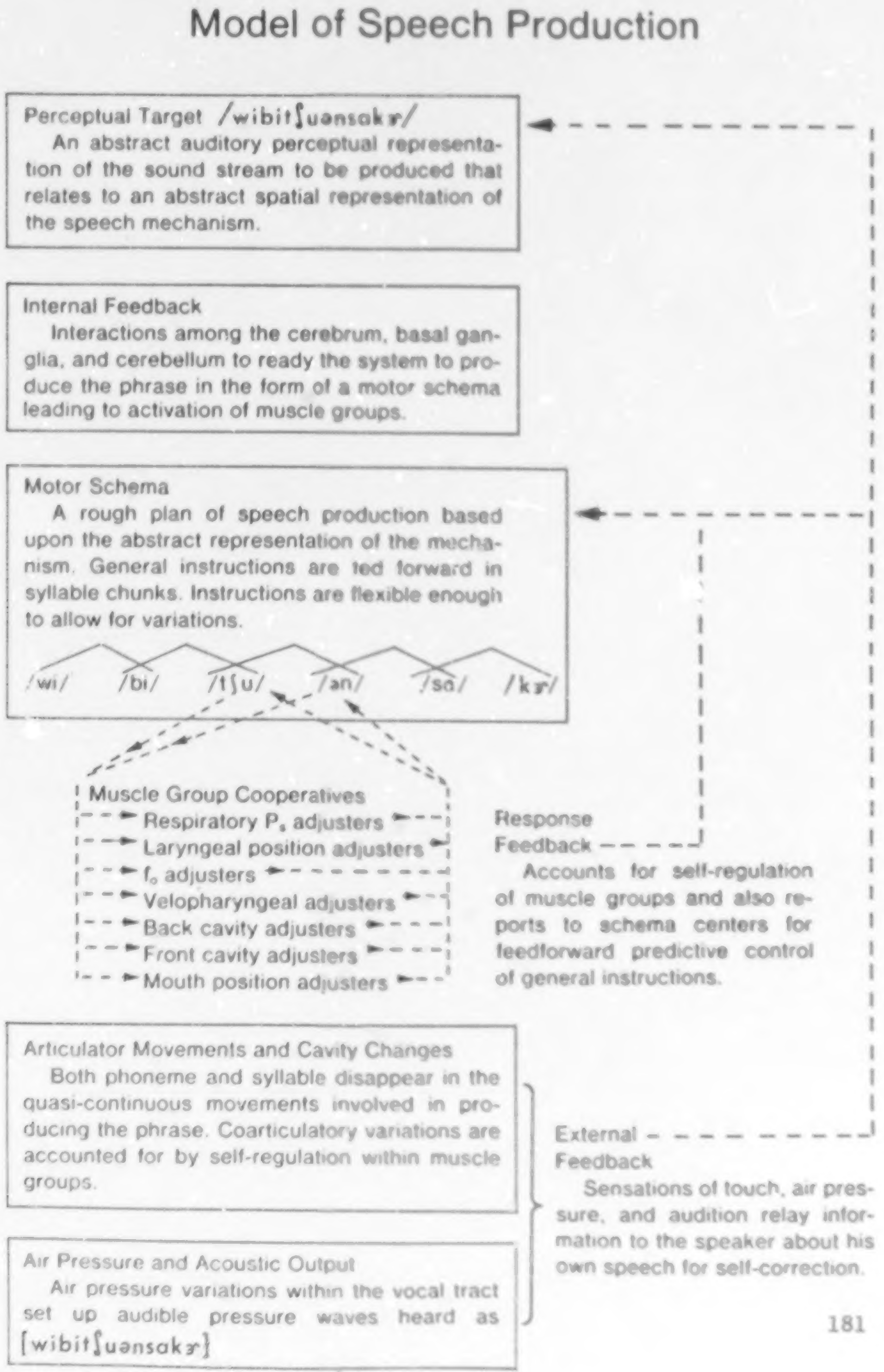
In this chapter, the goal is not to provide a comprehensive review of the literature on motor control, but rather to state a current view of how speech control may operate in skilled and nonskilled speakers--and to include examples from recent research in support of this view. Before discussing the ways in which feedback may operate during speech acquisition and during established speech, a brief consideration of the control mechanisms themselves and the experimental effects of altering the information they provide is in order.

CONTROL MECHANISMS FOR SPEECH

Normal speakers obtain feedback on their speech performance at a minimum of three levels of motor organization. Borrowing terms from Evarts (1971), writing on limb control--and applying those terms to speech--the levels are: internal feedback, response feedback, and knowledge of results or external feedback (Borden, 1979). Internal feedback is a network for information exchange entirely within the brain. The circuit includes the basal ganglia of the midbrain, the motor centers of the cerebrum, and the coordinative centers of the cerebellum. Response feedback is information from the joints, tendons, and especially muscles providing position and movement sense. This sensation, often termed 'proprioception' after Sherrington (1906) or 'kinesthesia,' a term that connotes awareness of the proprioception, arises as a response from the motor activity itself. The third level, external feedback, is information based upon the results of the motor patterns, and not upon the motor patterns themselves. Knowledge of results as applied to speech would include auditory and tactile information. The air- and bone-conducted sounds of speech are available to the speaker as are sensations of touch and of air pressure changes (Stevens & Perkell, 1977).

Figure 1 illustrates the multi-level control of speech suggested. In this model of speech production (Borden & Harris, 1980) the speaker is skilled and thus unconsciously knows the general sound of the phrase to be spoken, "We beat you in soccer," as well as the general requirements of the speech mechanism. This knowledge, labeled Perceptual Target in the figure, is translated into an abstract motor plan or schema by the interaction of the motor cortex, cerebellum, and basal ganglia of the brain. Neural activity in these areas has been recorded approximately 100 msec before speech, and although the purpose of the activity is not known, it can be postulated that the perceptual target is being translated into appropriate motor programs. Since the general schema of appropriate motor plans should be well known to the skilled speaker, the cerebellum can cooperate with the midbrain and cerebrum to apply well-practiced sequences of coordinated activities. An internal feedback loop could enable these centers to receive information from one another. The pre-speech activities represented in the top three boxes of the figure are considered to overlap in time, as do those represented in the two boxes at the bottom of the figure. As the motor schema develops, it is implemented sequentially. Kimura (1977) suggests, on the basis of evidence from patients with unilateral cerebral lesions, that the left hemisphere usually dominates in the control of the transitions within sequences of

Figure 1. A conceptualization of a skilled speaker's production of "We beat you in soccer." Reprinted with permission from Borden, G. J. & Harris, K. S. *Speech science primer*. Baltimore: Williams & Wilkins, 1980.



Model of Speech Production

Perceptual Target /wibitʃuənsakɚ/

An abstract auditory perceptual representation of the sound stream to be produced that relates to an abstract spatial representation of the speech mechanism.

Internal Feedback

Interactions among the cerebrum, basal ganglia, and cerebellum to ready the system to produce the phrase in the form of a motor schema leading to activation of muscle groups.

Motor Schema

A rough plan of speech production based upon the abstract representation of the mechanism. General instructions are fed forward in syllable chunks. Instructions are flexible enough to allow for variations.

Muscle Group Cooperatives

- Respiratory P_s adjusters
- Laryngeal position adjusters
- f_o adjusters
- Velopharyngeal adjusters
- Back cavity adjusters
- Front cavity adjusters
- Mouth position adjusters

Response Feedback

Accounts for self-regulation of muscle groups and also reports to schema centers for feedforward predictive control of general instructions.

Articulator Movements and Cavity Changes

Both phoneme and syllable disappear in the quasi-continuous movements involved in producing the phrase. Coarticulatory variations are accounted for by self-regulation within muscle groups.

Air Pressure and Acoustic Output

Air pressure variations within the vocal tract set up audible pressure waves heard as [wibitʃuənsakɚ]

External Feedback

Sensations of touch, air pressure, and audition relay information to the speaker about his own speech for self-correction.

181

different motor gestures. Thus, the left hemisphere may play a major role in switching from one muscle group to another.

Many groups of muscles cooperate to control the systems important for vocal tract shaping and for sound production. We know that the speech production system compensates easily for any constraint put upon it--whether the constraint is imposed from without, as in MacNeilage's (1970) example of speaking with a pipe clenched between the teeth, or is imposed by the speech context, as in moving to a /t/ from an open vowel rather than from a close vowel. The constant compensation for change that is evident in speech means either that muscle groups can autonomously regulate themselves, acting together as a 'coordinate structure' (Fowler, Rubin, Remez, & Turvey, in press), or that the muscle groups report on their performance to higher centers so that the motor schema can be altered appropriately. This information on the performance of the muscles, called response feedback, is relayed for speech primarily by muscle spindles embedded in the respiratory, laryngeal, and articulatory muscles. Response feedback from muscles is relatively rapid, as it precedes any movements that may result. Specialized muscle fibers, the spindles, lying in parallel with the main muscle fibers are thought to report information on muscle length and the rate at which the length is changing (Matthews, 1964). Simultaneous reports from a group of cooperating muscles could thus result in information on the relative contribution of each muscle in the group, and adjustments could be made for any imbalance. Constant responsiveness of the spindles is made possible by the simultaneous contraction of the main muscle and the spindle fibers during voluntary movement (Vallbo, 1971). Since the muscle spindles of the tongue lie in a three-directional pattern, they can provide complex three-dimensional information from many muscles simultaneously, yielding information on subtle changes in position and shaping. Cooper (1953) reported spindles to be especially prominent near the midline of the superior longitudinal muscle in a region proximal to the tip, in the most flexible part of the tongue.

When the muscle group cooperatives act to produce movements and changes in vocal cavity shape, the resulting changes in air pressure are made audible and form the sounds of speech. The speaker hears and feels his own speech being produced, and these sensations of touch and audition can be used for self-correction of errors. Auditory and tactile feedback can be classified as external feedback because they arise after the motor patterns for speech have been initiated. For the ballistic, fast-acting gestures of speech, such as the stop consonants, touch and audition occur too late to provide ongoing control of the motor commands. For continuants, however, they may serve to provide finer tuning of the productions to better match the perceptual targets. Tactile feedback is monitored by surface receptors responsive to light pressure changes and by subcutaneous receptors responsive to deep pressure. The anterior superior surface of the tongue is more sensitive to touch than the back or inferior parts of the tongue. The lips and the alveolar ridge of the palate are also sensitive to light touch (Ruch, 1960; Ringel & Ewanowski, 1965) and are important contact sites in speech. For auditory feedback, sound is conducted by air through the ear to the cochlea and also by bone through the skull of the speaker to the cochlea. The bone-conducted sound consists primarily of the low frequencies represented in the speech spectrum, while the air-conducted sound includes both high and low frequencies. The air- and bone-conducted sounds are approximately equal, however, in intensity (von Békésy, 1949).

EFFECTS OF ALTERED FEEDBACK

During the last thirty years, there have been many studies of normal speakers talking under various conditions of altered feedback. The auditory feedback of the speech signal has been delayed, amplified, attenuated, filtered, and masked altogether. The sense of touch has been diminished by blocking sensation from tongue and palatal receptors by the injection of anesthesia into the appropriate sensory nerves. The normal proprioceptive sense has been indirectly altered by changing the shape of the vocal tract with palatal prostheses or by imposing external constraints upon movement.

In all of these adverse feedback conditions, the result has been that speakers invariably compensate for any perturbation. When they hear their own speech delayed by a fraction of a second, they stall in an apparent attempt to let the auditory feedback catch up (Lee, 1950). When the auditory feedback is normal in timing but increased or decreased in intensity, speakers adjust their vocal intensity to try to match the feedback to their intended vocal effect (Lane & Tranel, 1971; Siegel & Pick, 1974). When the speech is filtered, speakers attempt to restore the missing frequencies (Garber & Moller, 1979). The adjustments made by speakers indicate that people are using auditory feedback to monitor their speech in situations of distorted feedback. Whether speakers use audition to monitor themselves to the same degree when the feedback is normal, however, is not known. When unable to hear themselves because of high level noise masking air-conducted sound and a bone vibrator masking bone-conducted sound, speakers continue to be highly intelligible. However, they increase vocal intensity and prolong vowels in trying to normalize the situation.

When oral anesthesia is applied to reduce tactile feedback, speech suffers subtle articulatory distortions, especially prominent in /s/-stop clusters (Borden, Harris, & Oliver, 1973). The speech is remarkably intelligible, however. When auditory and tactile feedback are diminished at the same time, the resulting speech remains intelligible, but the prosodic effects of the auditory masking are added to the articulatory effects of the tactile deprivation. The compensatory strategies that may operate to maintain intelligibility under these circumstances are difficult to measure at present. There is evidence of general muscle reorganization (Leanderson & Persson, 1972; Borden, Harris, & Catena, 1973), evidence of tongue retraction (Scott & Ringel, 1971), and evidence that airflow and air pressure are increased (Hutchinson & Putnam, 1974; Prosek & House, 1975).

It is perhaps easier to observe the compensatory mechanisms at work in conditions in which a bite block is inserted between the teeth, or in which there is a mechanical resistance applied to an articulatory movement, or, finally, in which the vocal tract is changed by inserting a false palate into the oral cavity. For vowel productions, speakers compensate immediately for a fixed jaw (Lindblom, Lubker, & Gay, 1979), apparently by altering tongue movement. When normal jaw movement is resisted mechanically, the lips compensate (Folkins & Abbs, 1975). Alterations of the vocal tract with prostheses change vocal tract coordinates and thus present conditions in which the subject has to recalibrate the system; thus, some trial and error is evident in developing compensatory patterns (Hamlet & Stone, 1976; Borden, Harris, Yoshioka, & Fitch, Note 1).

Another factor evident from altered feedback studies is related to the compensation detailed above. The fact that the feedback system for speech is highly redundant may account for much of the compensation observed. The system exists on so many levels that it is impossible to eliminate speech feedback altogether. When one channel is blocked, the speaker depends more upon another channel of information. Even in the rare case of adventitious auditory agnosia, the speaker can talk intelligibly despite the fact that all incoming sounds lack significance (Lassen, Note 2). Although the sound of his own speech is presumably useless to him, such a patient can rely upon proprioceptive information to monitor his performance.

How does one explain studies of animal movement in which even the proprioception is eliminated by severing the sensory roots providing all afferent information from movement? The animal subjects can continue, without visual feedback, to perform learned limb or jaw movements after complete deafferentation (Taub & Berman, 1968; Goodwin & Luschei, 1974; Polit & Bizzi, 1978). Perhaps the answer lies in the redundancy of the system. The internal feedback loop may be sufficient by itself. The brain knows that its performance is matching its intention so that no report from the periphery is required.

A final observation on the altered feedback experiments is the variability in the effects upon subjects. Delayed auditory feedback renders some subjects almost unable to speak, while others can fairly successfully ignore the auditory signal. Among subjects given a presumably identical injection of anesthesia to block sensation from the lingual nerve, there will be a noticeable effect on the speech of some, although there will be no effect that can be perceived on the speech of others (Borden, Harris, & Oliver, 1973). The degree to which this variability may reflect an equally large variability in the methods used to monitor speech under normal conditions can only be inferred.

FEEDBACK DURING SPEECH ACQUISITION

The critical time for the development of auditory, tactile, and proprioceptive feedback associations is regarded by many to be during the period of babbling. Before babbling, the infant makes vegetative, reflexive sounds connected with comfort, discomfort, and hunger. When babbling appears, it is mixed in with cooing but distinguished by its syllable-like repetitions of constricted vocal tract, consonant-like sounds releasing into more open vocal tract, vowel-like sounds. Babbling seems to be preprogrammed in the developmental sequence of motor activities, because both deaf babies and hearing babies babble in much the same way. After babbling for some time, the hearing baby presumably begins to attend to it and exert voluntary control over it. The baby hears and feels its own sound productions, which become increasingly similar to the language of its caretakers (Cruttenden, 1970; Weir, 1966). This requires the development of associations among the various feedback channels and motor patterns. In contrast, the deaf baby gradually babbles less (Mavilja, 1969, p. 151) as there is presumably less reward in vocal play; he can neither hear himself nor others.

Normal babies are born with an auditory mechanism that can distinguish between sounds that will, in many cases, be useful for phonemic distinctions later on. The work of Eimas, Siqueland, Jusczyk, and Vigorito (1971) and other studies in infant perception have shown that infants only a few weeks old are able to make fine auditory distinctions that correspond to contrasts of manner, place, and voicing in many languages. (See Kuhl, 1978, for a review.)

As an example of this work, we can take studies of the distinction between the sounds /r/ and /l/. Babies 2 and 3 months old were tested on speech-like stimuli varying in frequency change of the third formant, the most prominent acoustic cue used to distinguish /r/ from /l/. The sounds were presented contingent on the baby's sucking a pacifier. (With repetition of what the baby perceives as the same stimulus, sucking rate gradually decreases, but if a sound is introduced that the baby perceives as novel, sucking rate increases. Thus, a change in sucking rate indicates discrimination between habituated and novel sounds.) Infant discriminations were tested between stimuli that adults hear as /r/ and /l/, between stimuli that differ acoustically but are both heard as /r/ by adults, between stimuli also differing in F_3 change but heard by adults as /l/, and between identical stimuli. Infants gave evidence of a reliable increase in sucking rate when the stimuli paired were those heard by adults to be /r/ and /l/ as opposed to the other pairs (Eimas, 1975). So, babies seem to hear /r/-/l/ differences, among many others, and in their early babbling /r/- and /l/-like sounds are frequent (Clark & Clark, 1977); but during the second, more volitional stage of babbling, /r/ and /l/ drop out (Jakobson, 1968). The increasing use of voluntary control in babbling, although important in developing motor control, thus produces some interesting changes in the vocal repertoire.

As babbling interweaves with the beginnings of first words, another aspect of production and perception is added. Not only has the child assumed some voluntary and increasingly precise control over sound production, but these sounds have started to take on meanings. In the production of meaningful utterances /r/ and /l/ appear late (Templin, 1957; Jakobson, 1968). Furthermore, when embedded in a linguistic task, perception of /r/ and /l/ is also late (Shvachkin, 1973). It seems that acoustic features important to language are detected soon after birth, but speech perception in context takes time to develop (Edwards, 1974; Zlatin & Koenigsknecht, 1975). It may also be the case that feedback sufficient to gain control over relatively simple oral sound productions develops early, but the feedback mechanisms and motor control necessary to produce complex speech patterns in which semantic processes are involved develop slowly. Evidence of this slowly developing control comes from recordings of children practicing their newly discovered language by themselves. The repetitions and variations of phrases resemble phonetic practice as much as they do vocabulary practice (Weir, 1966, pp. 166-167). In this sample, Weir's child was practicing variations in the initial consonant:

(1) fumbelina (2X)
(2) tumbelina
(3) lumbelina
(4) Thumbelina (2X)

In developing control of speech, children seem to use adult models as their perceptual targets, but their perception of the targets is apt to be rather undifferentiated and their ability to reproduce the targets imprecise and variable. In general, their perceptual ability seems to develop faster than their production ability. In identification and reproduction tasks on sets of synthetic CVC syllables in continua from 'light' to 'white,' from 'light' to 'write,' and from 'white' to 'write,' four-year-old children could identify the syllables better than they could repeat them (Menyuk & Anderson, 1969).

Strange and Broen (in press) compared perception and production of /r/-/l/ along with other semivowels in three-year-old children and found that they were good at differentiating such contrasts as 'rake' and 'lake,' and that they perceived the phoneme boundary in a synthetic series much as college students do. Their most interesting finding was that children who have mastered the /r/-/l/ contrast in their speech, or who showed few distortions, were good at identifying clear examples of the contrast--whether delivered by live voice, by tape-recorded voice, or synthetically produced; while of the children with many distortions and substitutions in their speech, half made very few identification errors but half had some difficulty perceiving the contrasts. There does seem to be some relationship, then, between perception and production, but even in the four worst cases, perception was better than production.

The relationship between a child's perception and production is complex. It changes in mysterious ways as the child develops. (Adele Gerber of Temple University illustrates this with a language sample from a child named Eric (Note 3). At 14 months, Eric called a dog 'fa fa.' His mother would say, "See the dog? Ruff-ruff. That's a ruffy dog." Eric would say "fa-fa." By 19 months, he was saying 'goggy' when he pointed to a dog. His grandmother said to him, "You used to call him a 'fa fa'," whereupon Eric said, "Ruffy gog.")

Although in advance of production, perception does seem to be developmental. Chaney and Menyuk (Note 4) reported that four-year-olds who produced [w] for /r/ and /l/ and six-year-olds who produced /r/ errors were not as accurate as controls with good articulation in pointing to pictures of 'light,' 'write,' or 'white' when adults produced them; further, the four-year olds could only identify between 33% and 46% of their own tape-recorded productions.

This question of how children perceive themselves during their parallel refinement of perception and production is of interest to those theorizing on feedback and motor control (see McReynolds, 1978, for a review). In the aforementioned Chaney and Menyuk study, children heard themselves on a tape recording and failed to point to the picture they had previously named. Also, Locke and Kutz (1975) found that five-year-old children who say 'wing' for 'ring' can perceive the /r/ correctly in adult speech but when they hear their own tape-recorded misarticulations of 'ring,' they are more likely to point to a picture of a 'wing' than a 'ring.'

On the other hand, some people report that children may be hearing some difference in their own misarticulations that adults fail to hear. Support for this is largely anecdotal. An example:

194

Child: She's wearing a wing.
Adult: A wing?
Child: No, not a bird wing, a wedding wing.

Kornfeld (1971) found a spectrographic difference between the [w] in a child's production of [gwæs] for 'glass' and the [w] in the same child's production of [gwæs] for 'grass.' This suggests that there may be some differences in production that are not phonemically significant to adults.

The apparent discrepancy between the notion that children are perceiving a difference that adults fail to perceive and the view that children usually perceive much as adults do, but lack production proficiency, may be partly resolved if we realize that the anecdotal support for the first notion comes from children making discriminations of their own utterances while speaking, but support for the second idea comes from children's discrimination of others' speaking or of their own tape-recorded speech. During live speech, children can feel themselves as well as hear themselves. Thus, they may sense small differences between 'wing' for bird and wedding 'wing' that neither they nor adults perceive as different by audition alone. One could infer either that the children were aware of a difference in intention or that there was some phonemically insignificant difference in muscle activity that they could sense. Spectrographic differences support the second alternative. The logic of this explanation might lead us to the conclusion that in these cases, children are using proprioception more than audition to monitor themselves, or, more conservatively, that muscle sense is needed in addition to the hearing sense for some children to distinguish their own /w/-/w^{r}/ distinctions.

Development of feedback mechanisms has been difficult to study. Studies of DAF (Delayed Auditory Feedback) on the speech of children have shown older children (7 to 9 years) to be more affected by the delay than younger children (4 to 6 years) (Chase, Sutton, First, & Zubin, 1961), but later studies found younger children to be more affected than older children or adults (MacKay, 1968). Using DAF to test the development of auditory feedback may not be the best method--the technique forces attention to audition and creates an artificial discrepancy between muscle information and the auditory information returned to the speaker (Borden, Dorman, Freeman, & Raphael, 1976). Studies of children speaking in noise, while hearing their own voices amplified, show that the children alter their vocal intensity appropriately but compensate to a different degree than do adults. The children reduced vocal intensity less under amplified feedback than did adults, but did not differ significantly in the degree to which they increased vocal intensity when speaking in noise (Siegel, Pick, Olsen, & Sawin, 1976). By four years of age children compensate much like adults for any artificial interference in feedback. Nerve-block anesthesia of the tongue resulted in the same proportion of affected articulation in four-year-olds as was found for adults (Borden, 1976). The redundancy of the feedback systems and the inaccessibility of the proprioceptive system to intervention prevent a direct comparison of the different feedback systems. A predictive feedforward system with CNS (Central Nervous System) control may be fairly well developed, however, by four years. Also, 'muscle sense' may take over when the auditory or tactile senses are diminished. Happily, these problems for the researcher are assets for the speaker.

We are left with the realization that we know very little about the development and use of feedback control mechanisms in children's learning to speak. By the time children are old enough to be tested, they are fairly skilled speakers despite the fact that some speech refinements are still being acquired. We know, however, that children are much more competent than adults in learning new languages. There is an apparent plasticity of the neurological networks that maintains its flexibility until approximately the age of puberty (Lenneberg, 1967; Penfield & Roberts, 1959). It behooves us to test children on unfamiliar or more recently learned items as well as on well-known utterances--and to compare young children, older children, and adults on their relative dependence upon feedback when producing speech gestures varying in familiarity.

CONTROL OF ESTABLISHED SPEECH

There have been some studies of altered feedback on adults in which the effects were studied on the speaker's native language, a second language, and an entirely novel language. One such study that was well controlled and designed (MacKay, 1970) tested 21 English-German bilinguals, some for whom German was the native language and others for whom English was the native language. Subjects read 15-syllable sentences in both languages and in a completely unfamiliar language, Congolese. In the experimental condition, they read the sentences under delayed auditory feedback. Native speakers of German made more 'stuttering-like responses' when speaking English than when speaking German, and native speakers of English made more 'stuttering-like responses' when speaking German. Both groups of speakers had more difficulty with the unfamiliar language, Congolese. DAF was thus more of an interference to the less familiar language and most disruptive of the novel Congolese sentences.

In an attempt to control the variable of attention, MacKay had the subjects read in German, and in English, while listening to their own voices speaking the opposite language. They spoke both native and second languages more slowly with this distraction, but it interfered more with their native language. MacKay suggests that this may have been caused by increased concentration on the articulation of the less familiar language. A slight variant of this explanation might be that the native language could be put on automatic control to a degree so that subjects could do two things at once: speak and listen to an irrelevant voice--while for the second language they had to maintain more control and could not, at the same time, attend to an unrelated voice.

Why, then, would DAF interfere more with the less familiar language? Would not MacKay's suggestion hold here as well? To me, the important variables are the intensity and nature of the auditory information. In the DAF condition, the distorted signal was delivered at a 95 dB sound pressure level and the speaker recognized the signal as his own ongoing speech, delayed. The irrelevant voice condition was not amplified to such an extent and even if recognized by the speaker, was not associated with the ongoing speech. Thus, when the speaker is most skilled, as in speaking his or her native language, less control is needed and there is less interference from a distortion of feedback; but for the same reason, the speaker is free is attend to something unrelated to the speech.

One barrier to an interpretation of DAF studies has been that DAF is described as interference with auditory feedback. The crux of the DAF effect is not just that the auditory feedback is delayed, but that the auditory signal is amplified beyond ignoring and presents competing information with the proprioceptive feedback within the speaker. Delayed auditory feedback presents the subject with an auditory-proprioceptive mismatch.

We get a glimpse of how speech may be controlled by skilled speakers and how perception and production may interact when we study speakers learning a new language. In a study currently underway at Haskins Laboratories, we are finding that skilled speakers produce unfamiliar gestures in accordance with their perceptions of them. One speaker will perceive the sounds as novel and use trial and error strategies to match the unfamiliar target; another speaker will also perceive the sounds as novel but will attempt to incorporate any new sound into his own production system and make unfamiliar items familiar in terms of motor patterns. The difference in the way subjects link the perceptual target with the production target may determine the difference between utterances by speakers imitating syllables under various conditions of feedback deprivation. The speaker attempting to match a novel perceptual target, such as /χi/ shows much more variability on unfamiliar gestures than on familiar gestures, but the speaker adhering to familiar perceptual targets, perceiving /χi/ as a variant of /ʃi/ or /hi/, evidences variability under various feedback conditions but no more variability on unfamiliar than on familiar gestures. This investigation shows, too, that even for skilled speakers producing familiar speech sounds, feedback plays a part in the fine tuning of the match between the actual production and the perceptual target. For example, the second formant of vowels was found to be higher when subjects were able to hear themselves, whether the vowel was the familiar /i/ or the less familir /y/ (Borden et al., Note 1).

Perception thus has a strong influence on production. It is also true that production, or language experience, affects perception of speech sounds. When listeners are presented with a series of synthetic sounds that form an acoustic continuum between two phonemes, each member of the series differing from the next by an equal acoustic change, there is a strong tendency for the sounds to be heard categorically. That is, some members of the series are heard as one phoneme, some as the other phoneme, and discrimination between members of the series is high at the phoneme boundary but low among the sounds labeled as a particular phoneme (Liberman, Harris, Hoffman, & Griffith, 1957).

It has been demonstrated that linguistic experience as well as psychoacoustic ability affects adult perception of phonemic boundaries along such acoustic continua. The positive influence of production upon perception has been evident in an increased ability to discriminate acoustic dimensions related to one's own linguistic experience, and a corresponding decreased ability if the distinction is not related to one's linguistic experience (see Strange & Jenkins, 1978, for a review). For example, one study (Miyawaki, Strange, Verbrugge, Liberman, Jenkins, & Fujimura, 1975) compared speakers of Japanese with speakers of American English on their perception of the liquids /r/ and /l/. These sounds are contrastive in English, as in 'rake' and 'lake,' but not contrastive in Japanese. Thus, American English listeners divided a series of synthetic syllables varying in F_3 starting frequency into two categories: those that sounded like /ra/ and those that sounded more like

/la/. Listeners from Japan tended to hear all the stimuli as /ra/. To test discrimination, listeners were presented with a series of three items from the set--two identical and one different. Subjects indicated whether the different syllable was heard first, second, or third. American listeners could tell the difference 80% of the time if the compared stimuli were near the /r/-/l/ boundary, but their discrimination scores fell to between 40-60% if the stimuli were those that they identified as a single phoneme. The Japanese listeners, in general, showed no sharp discrimination peak. In discrimination of non-speech stimuli consisting of F_3 without the other formants, Japanese and American groups both discriminated betwen 66% and 89% correct. It appears that they report auditory differences among non-speech sounds equally well, but that when listening to speech-like sounds they report only the differences important in their own speech.

The Miyawaki study used as Japanese subjects, students living in Japan. They had studied English in school but they had had little practical experience in speaking English. Another study of a Japanese group and an American group of speakers living in Japan pointed to a more complex relationship between perception and production (Goto, 1971). Both groups were recorded producing words containing /r/ or /l/, such as 'collect,' 'correct,' 'play,' and 'pray.' Then each subject was asked to identify either /r/ or /l/ in the words as recorded by themselves and by the others. American speakers were good at identification of /r/ and /l/ in their own speech and in the speech of Japanese speakers. Some of the Japanese subjects produced the contrast to the satisfaction of American listeners and some did not, but the Japanese listeners were poor at identification, whether they could produce the contrast or not. Goto concluded that the Japanese speakers good at production but poor in identification must be using kinesthetic feedback. Some of the Japanese subjects were further tested on a combined identification-discrimination task. They responded 'r-l different' or 'l-l same' to pairs of words, 'pray-play' or 'collect-collect,' as recorded by American speakers. Inspection of the data presented indicates that discrimination was more difficult for the Japanese than identification. Goto tested himself on the identification task after one month of studying English conversation, after four months, and after ten months. There was not much improvement in his perception, but no mention was made of whether his speech improved during that time. The Goto study has perplexed many investigators because it seems to indicate that correct production of a phonemic contrast can be developed in the absence of any auditory perception of the difference. It may be, however, that despite poor discrimination, some degree of auditory identification is necessary initially to associate the new production with some sound of the speaker's first language. Both the production and perception of a novel phoneme in the second language may be related to a variation of an item in the first language. It follows that whatever feedback and feedforward mechanisms operate to control the sound patterns of the first language are quickly adopted for the second. It would be interesting to track the relative importance of feedback information as skilled speakers produce their native speech patterns and as they learn new patterns.

CONCLUSION

We have reviewed evidence for at least three levels of information flow that can be used to direct the production of speech: internal feedback, response feedback, and external feedback. Internal feedback is made possible by circuits among the cerebrum, midbrain, and cerebellum of the brain. Response feedback from much of the speech musculature is relayed from muscle spindles, and, added to any information from tendon and joint receptors, is thought to form the complex sensation of movement and position. External feedback through the auditory and tactile systems yields knowledge of results to the speaker, knowledge that can be used for fine tuning of the speech signal or for correction of errors.

Experimental interference with external feedback has demonstrated the remarkable compensatory abilities of speakers in general, but has also pointed to large variation in effects among individual speakers.

As children acquire speech, they seem to build upon their innate acuity in detection of speech sounds, but perception of distinctions embedded in meaningful stimuli evidently develops with age along with, and somewhat in advance of their production abilities. Self-perception must play an important role in forming associations between speech perception and speech production in speakers acquiring new speech patterns. Skilled adult speakers perceive speech in accordance with their own linguistic experience, and are generally far more inhibited in producing new patterns than are children.

Learning a motor skill seems to require some knowledge of results (Adams, 1971), and children who engage in variable practice can adapt quickly to learn new tasks (Kelso & Norman, 1978). Children acquiring their first language engage in variable practice and seem to depend upon their well-developed auditory discrimination to refine their speech and to give them knowledge of their progress. The degree to which the other feedback mechanisms are used can only be imagined at present. During the critical period of speech learning (Lenneberg, 1967; Marler, 1975), children show a tendency toward experimentation and the active use of feedback. It does seem, however, that the perceptual representations and the articulation programs in young children are both unstable, that perception stabilizes ahead of motor control, and thus, older children are better able to benefit from feedback as they gain motor proficiency (Newell & Kennedy, 1978). Many of the references cited above refer to the development of motor learning in areas other than speech. We must study this development more closely in speech. Do we indeed depend less upon feedback as we become more skilled as speakers? Preliminary studies indicate that this may be the case.

The strategies used in learning a new phonetic system may depend upon whether the speaker is still within the critical period for language learning or well beyond it. There are indications that children learning a second language keep the feedback channels open. A study of Puerto Rican children learning English (Williams, 1974) showed perception to improve as experience with the new language increased, but the younger children changed more rapidly toward English perceptual boundaries than did older children. The increased sensitivity to the contrasts important to the new language was found to interfere temporarily with the native language.

Adults learning a new language, however, seem to base the new language on their native language. They use feedback, but, lacking established links between feedback and the new production programs, their new speech gestures tend to be modifications of their old system, complete with whatever degree of automaticity is involved in that system. Skilled speakers are good at perceiving distinctions important in their native language (Miyawaki et al., 1975) but poor at perceiving distinctions important in a less familiar language, even if they produce the distinctions passably well (Goto, 1971). Only to the degree that adults can ignore their previously learned sound system and can become "child-like" in their freedom to experiment and in their sensitivity to their own productions, will they enjoy success in achieving the suprasegmental and segmental nuances of a new language.

REFERENCE NOTES

1. Borden, G. J., Harris, K. S., Yoshioka, H., & Fitch, H. *Feedback and feedforward mechanisms used by speakers producing familiar and novel speech patterns*. Paper presented at the Ninth International Congress of Phonetic Sciences, Copenhagen, 1979.
2. Lassen, N. A. *The physiology and pathophysiology of language functions as illustrated by measurements of the regional blood flow in the cortex of the brain*. Special lecture presented at the Ninth International Congress of Phonetic Sciences, Copenhagen, 1979.
3. Gerber, A. Personal communication, 1979.
4. Chaney, C. F., & Menyuk, P. *Production and identification of /w,r,l,j/ in normal and articulation impaired children*. Paper presented at the American Speech and Hearing Association Convention, Washington, D.C., 1975.

REFERENCES

Adams, J. A. A closed-loop theory of motor learning. *Journal of Motor Behavior*, 1971, *3*, 111-150.

Békésy, G. von. The structure of the middle ear and the hearing of one's own voice by bone conduction. *Journal of the Acoustical Society of America*, 1949, *21*, 217-232.

Borden, G. J. The effect of mandibular nerve block upon the speech of four-year-old boys. *Language and Speech*, 1976, *19*, 173-178.

Borden, G. J. An interpretation of research on feedback interruption in speech. *Brain and Language*, 1979, *7*, 307-319.

Borden, G. J., Dorman, M. F., Freeman, F. J., & Raphael, L. J. Electromyographic changes with delayed auditory feedback of speech. *Journal of Phonetics*, 1976, *4*, 1-8.

Borden, G. J., & Harris, K. S. *Speech science primer*. Baltimore, Md.: Williams & Wilkins, 1980, 155-156.

Borden, G. J., Harris, K. S., & Catena, L. Oral feedback II. An electromyographic study of speech under nerve-block anesthesia. *Journal of Phonetics*, 1973, *1*, 297-308.

Borden, G. J., Harris, K. S., & Oliver, W. Oral feedback I. Variability of the effect of nerve-block anesthesia upon speech. *Journal of Phonetics*, 1973, *1*, 289-295.

Chase, R. A., Sutton, S., First, D., & Zubin, J. A developmental study of changes in behavior under delayed auditory feedback. *Journal of Genetic Psycholgy*, 1961, *99*, 101-112.

Clark, H. H., & Clark, E. V. *Psychology and language*. New York: Harcourt Brace Jovanovich, 1977.

Cooper, S. Muscle spindles in the intrinsic muscles of the human tongue. *Journal of Physiology*, 1953, *122*, 193-202.

Cruttenden, A. A phonetic study of babbling. *British Journal of Disorders of Communication*, 1970, *5*, 110-117.

Edwards, M. L. Perception and production in child phonology: The testing of four hypotheses. *Journal of Child Language*, 1974, *1*, 205-219.

Eimas, P. D. Auditory and phonetic coding of the cues for speech: Discrimination of the r-l distinction by young infants. *Perception & Psychophysics*, 1975, *18*, 341-347.

Eimas, P. D., Siqueland, E. R., Jusczyk, P., & Vigorito, J. Speech perception in infants. *Science*, 1971, *171*, 303-306.

Evarts, E. V. Feedback and corollary discharge: A merging of the concepts. In *Central control of movement*. *Neurosciences Research Program Bulletin*, 1971, *9*, 86-112.

Folkins, J. W., & Abbs, J. H. Lip and jaw motor control during speech: Responses to resistive loading of the jaw. *Journal of Speech and Hearing Research*, 1975, *18*, 207-220.

Fowler, C. A., Rubin, P., Remez, R. E., & Turvey, M. T. Implications for speech production of a general theory of action. In B. Butterworth (Ed.), *Language production*. New York: Academic Press, in press.

Fry, D. B. The control of speech and voice. In H. Kalmus (Ed.), *Regulation and control in living systems*. New York: John Wiley & Sons, 1966, 137-154.

Garber, S. R., & Moller, K. T. The effects of feedback filtering on nasalization in normal and hypernasal speakers. *Journal of Speech and Hearing Research*, 1979, *22*, 321-333.

Goodwin, G. M., & Luschei, E. S. Effects of destroying the spindle afferents from the jaw muscles upon mastication in monkeys. *Journal of Neurophysiology*, 1974, *37*, 967-981.

Goto, H. Auditory perception by normal Japanese adults of the sounds "L" and "R." *Neuropsychologia*, 1971, *9*, 317-323.

Hamlet, S. L., & Stone, M. Compensatory vowel characteristics resulting from the presence of different types of experimental dental prostheses. *Journal of Phonetics*, 1976, *4*, 199-218.

Hutchinson, J. M., & Putnam, A. H. B. Aerodynamic aspect of sensory deprived speech. *Journal of the Acoustical Society of America*, 1974, *56*, 1612-1617.

Jakobson, R. *Child language, aphasia, and phonological universals*. The Hague: Mouton, 1968.

Kelso, J. A. S., & Norman, P. E. Motor schema formation in children. *Developmental Psychology* 1978, *14*, 153-156.

Kimura, D. Acquisition of a motor skill after left-hemisphere damage. *Brain*, 1977, *100*, 527-542.

Kornfeld, J. Theoretical issues in child phonology. In *Proceedings of the Seventh Annual Meeting of the Chicago Linguistic Society* (University of Chicago), 1971, *CLS 7*, 454-468.

Kuhl, P. K. Predispositions for the perception of speech-sound categories: A species-specific phenomenon? In F. D. Minifie & L. L. Lloyd (Eds.), *Communicative and cognitive abilities*. *Early behavioral assessment*. (NICHD Series). Baltimore, Md.: University Park Press, 1978, 229-255.

Lane, H. L., & Tranel, B. The Lombard sign and the role of hearing in speech. *Journal of Speech and Hearing Research*, 1971, *14*, 677-709.

Leanderson, R., & Persson, A. The effect of trigeminal nerve block on the articulatory EMG activity of facial muscles. Acta Otolaryngologica, 1972, 74, 271-278.

Lee, B. S. Effects of delayed speech feedback. Journal of the Acoustical Society of America, 1950, 22, 824-826.

Lenneberg, E. H. Biological foundations of language. New York: John Wiley & Sons, 1967.

Liberman, A. M., Harris, K. S., Hoffman, H. S., & Griffith, B. C. The discrimination of speech sounds within and across phoneme boundaries. Journal of Experimental Psychology, 1957, 54, 358-368.

Lindblom, B., Lubker, J., & Gay, T. Formant frequencies of some fixed-mandible vowels and a model of speech motor programming by predictive simulation. Journal of Phonetics, 1979, 7, 147-161.

Locke, J. L., & Kutz, K. J. Memory for speech and speech for memory. Journal of Speech and Hearing Research, 1975, 18, 176-191.

MacKay, D. G. Metamorphosis of a critical interval: Age-linked changes in the delay in auditory feedback that produces maximal disruption of speech. Journal of the Acoustical Society of America, 1968, 43, 811-821.

MacKay, D. G. How does language familiarity influence stuttering under delayed auditory feedback? Perceptual and Motor Skills, 1970, 30, 655-669.

MacNeilage, P. F. Motor control of serial ordering of speech. Psychological Review, 1970, 77, 182-196.

Marler, P. On the origin of speech from animal sounds. In J. F. Kavanagh & J. E. Cutting (Eds.), The role of speech in language. Cambridge, Mass.: MIT Press, 1975, 11-37.

Matthews, P. B. C. Muscle spindles and their motor control. Physiological Review, 1964, 44, 219-288.

Mavilja, M. P. Spontaneous vocalization and babbling in hearing impaired infants. Unpublished doctoral dissertation, Columbia University, 1969.

McReynolds, L. V. Behavioral and linguistic considerations in children's speech production. In J. F. Kavanagh & W. Strange (Eds.), Speech and language in the laboratory, school, and clinic. Cambridge, Mass.: MIT Press, 1978, 127-164.

Menyuk, P., & Anderson, S. Children's identification and reproduction of /w/, /r/, and /l/. Journal of Speech and Hearing Research, 1969, 12, 39-52.

Miyawaki, K., Strange, W., Verbrugge, R., Liberman, A. M., Jenkins, J. J., & Fujimura, O. An effect of linguistic experience: The discrimination of [r] and [l] by native speakers of Japanese and English. Perception & Psychophysics, 1975, 18, 331-340.

Newell, K. M., & Kennedy, J. A. Knowledge of results and children's motor learning. Developmental Psychology 1978, 14, 531-536.

Penfield, W., & Roberts, L. Speech and brain-mechanisms. Princeton, N.J.: Princeton University Press, 1959.

Polit, A., & Bizzi, E. Processes controlling arm movements in monkeys. Science, 1978, 201, 1235-1237.

Prosek, R. A., & House, A. S. Intraoral air pressure as a feedback cue in consonant production. Journal of Speech and Hearing Research, 1975, 18, 133-147.

Ringel, R. L., & Ewanowski, S. J. Oral perception: Two-point discrimination. Journal of Speech and Hearing Research, 1965, 8, 389-398.

Ruch, T. C., & Fulton, J. F. Medical physiology and biophysics. Philadelphia: Saunders Co., 1960.

Scott, C. M., & Ringel, R. L. Articulation without oral sensory control.

Journal of Speech and Hearing Research, 1971, 14, 804-818.

Sherrington, C. S. The integrative action of the nervous system. London: Constable, 1906.

Shvachkin, N. K. The development of phonemic speech perception in early childhood. In C. Ferguson & D. Slobin (Eds.), Studies of child language development. New York: Holt, Rinehart, and Winston, 1973, 91-127.

Siegel, G. M., & Pick, H. L., Jr. Auditory feedback in the regulation of the voice. Journal of the Acoustical Society of America, 1974, 56, 1618-1624.

Siegel, G. M., Pick, H. L., Jr., Olsen, M. G., & Sawin, L. Auditory feedback in the regulation of vocal intensity of preschool children. Developmental Psychology, 1976, 12, 255-261.

Stevens, K. N., & Perkell, J. S. Speech physiology and phonetic features. In M. Sawashima & F. S. Cooper (Eds.), Dynamic aspects of speech production. Tokyo: University of Tokyo Press, 1977, 323-341.

Strange, W., & Broen, P. A. Perception and production of approximant consonants by three-year-olds: A first study. In G. Yeni-Komshian, J. F. Kavanagh, & C. A. Ferguson (Eds.), Child phonology: Perception, production and deviation. Proceedings of an NICHHD conference, Bethesda, Md., May 1978 (in press).

Strange, W., & Jenkins, J. J. Role of linguistic experience in the perception of speech. In R. D. Walk & H. L. Pick, Jr. (Eds.), Perception and experience. New York: Plenum Press, 1978, 125-169.

Taub, E., & Berman, A. J. Movement and learning in the absence of sensory feedback. In S. J. Freeman (Ed.), The neuropsychology of spatially oriented behavior. Homewood, Ill.: Dorsey Press, 1968, 173-192.

Templin, M. Certain language skills of children. Minneapolis: University of Minnesota Press, 1957.

Vallbo, A. B. Muscle spindle response at the onset of isometric voluntary contractions in man. Journal of Physiology, 1971, 318, 405-431.

Weir, R. H. Some questions on the child's learning of phonology. In F. Smith & G. A. Miller (Eds.), The genesis of language. Cambridge, Mass.: MIT Press, 1966, 153-168.

Williams, L. Speech perception and production as a function of exposure to a second language. Unpublished doctoral dissertation, Harvard University, 1974.

Zlatin, M. A., & Koenigsknecht, R. A. Development of the voicing contrast: Perception of stop consonants. Journal of Speech and Hearing Research, 1975, 18, 541-553.

BRAIN LATERALIZATION IN 2-, 3-, AND 4-MONTH-OLDS FOR PHONETIC AND MUSICAL TIMBRE DISCRIMINATIONS UNDER MEMORY LOAD[*]

Catherine T. Best, Harry Hoffman[+], and Bradley B. Glanville[++]

Abstract. Eight male and eight female infants of 2, 3, and 4 months of age each completed a four-part dichotic test for ear differences in short term memory-based discriminations among synthetic syllable-initial stop consonants, and among synthesized renditions of different musical instruments playing the same-pitched note. A dichotic cardiac OR (orienting response) habituation/dishabituation procedure involving long ISIs and discrete stimulus presentations (Glanville, Best, & Levenson, 1977) was used. Results supported an REA (Right Ear Advantage) for speech discriminations, and an LEA (Left Ear Advantage) for musical timbre discriminations, in the 3- and 4-month-olds. The 2-month-olds, however, showed only the LEA for musical timbre discriminations, providing no reliable evidence of ability to discriminate syllable-initial stop consonants with *either* hemisphere under memory load. Of the entire sample considered as individuals, 33/48 showed some evidence of music discrimination, with an LEA for 22/33; also, 33/48 infants showed some evidence of speech discrimination, with an REA in 24/33. Implications for theories about the development of brain lateralization, and of general perceptual differences for the two classes of acoustically complex auditory stimuli, are discussed. The possibility of using the procedure with individuals and atypical populations is also addressed.

---

[*] Parts of these data were presented at the 1st *International Conference on Infant Studies*, in Providence, Rhode Island (March, 1978).

[+] North Dakota State University, Fargo

[++] California State University, Chico

Acknowledgment: The completion of this work was partially supported by NINCDS Postdoctoral Fellowship Grant #NS5085 to the first author, and by grants to the third author from North Dakota State University, Fargo, for work conducted in the Infant Psychophysiology Laboratory. The authors are very grateful to Dr. David Pisoni of Indiana University for providing us with the speech stimuli for this study, and to the following people for thoughtful discussions of the results and helpful comments on earlier versions of this paper: Drs. Hiram E. Fitzgerald, Lauren J. Harris, Jerre Levy, Kristine MacKain, Michael Studdert-Kennedy, and Robert Bundy. We would also like to thank Patrick K. Ackles for advice on certain statistical and computer techniques. We are especially indebted to the infant participants' parents for their involvement and sincere interest.

[HASKINS LABORATORIES: Status Report on Speech Research SR-59/60 (1979)]

INTRODUCTION

In most right-handed human adults, the left cerebral hemisphere is specialized for language processing (about 85-95% -- Branch, Milner, & Rasmussen, 1964; Levy, 1974), as well as speech production, speech perception, and motoric and cognitive abilities related to language use, such as handwriting and serial order recall (e.g., Cohen, 1973; Natale, 1977), respectively. In dichotic listening studies, this left hemisphere advantage is reflected in superior right ear performance on tasks involving speech or speech-like auditory stimuli and/or language-related abilities (e.g., Cutting, 1974; Kimura, 1961; Kimura & Folb, 1968; Shankweiler & Studdert-Kennedy, 1967; Studdert-Kennedy & Shankweiler, 1970). Conversely, for most adults the right hemisphere excels at a range of visuo-spatial skills such as map-reading or mental paper-folding, and at other gestalt-like or holistic processes (e.g., Benton, 1962; Bogen, 1969; Carmon & Bechtoldt, 1969; Carmon & Benton, 1969; Fontenot & Benton, 1972; Hebb, 1939; Kimura, 1969; Levy, 1974; Levy-Agresti & Sperry, 1968; Nebes, 1970; Harris & Best, Note 1). The right hemisphere in most adults is superior to the left for recognition and recall of relatively complex nonlinguistic (nonphonetic) auditory information such as the melodic properties of both music and speech, and the quality of musical chords (e.g., Blumstein & Cooper, 1974; Deutsch, 1975; Gordon, 1970; Milner, 1962; Teuber, 1970). Therefore, dichotic listening studies using nonlinguistic auditory stimuli like those just described typically yield a left ear advantage (cf. Kimura, 1964; Shankweiler, 1966;--for a general review of functional hemispheric asymmetries, see Dimond & Beaumont, 1974).

The fact of cerebral dominance suggests there may be a potent, neurobiologically-based "linguistic/nonlinguistic" distinction (at least, very broadly defined) in human perception of auditory events. That is, the human nervous system is apparently organized for a functional auditory perceptual differentiation that is relevant to a dichotomy between the set of cognitive/perceptual responses typically involved in linguistic abilities, versus the set of responses appropriate for complex nonlinguistic auditory skills. This is a very striking characteristic, and may be uniquely human (e.g., Levy, 1974; Liberman, 1970) or at least uniquely related to the need for precise motor control in species-specific human and songbird vocal communication (e.g., Studdert-Kennedy, 1976).[1] The neural functional lateralization just described would be of obvious importance in human ontogeny, particularly in language acquisition, and indeed would have been crucial to our evolution as a species. Yet we know remarkably little about its ontogenetic and evolutionary basis--in other words, we still are largely in the dark about the nature of what behaviors are actually lateralized, and how they came to be lateralized in adult, present-day humans. The research reported in this paper focuses on the what, and may provide insight on the how, of auditory perceptual lateralization in early human ontogeny. Specifically, we found that during the first third of the first year of postnatal life there is at least some degree of lateralization in auditory perception and memory. Furthermore, there are early age differences in the manifestation of auditory lateralization involving short term memory. By implication, there may be important early age changes in the nature of the "linguistic/nonlinguistic" distinction in perception of auditory events.

The first and best-known model for the ontogeny of human cerebral lateralization was proposed by Lenneberg (1967). Lenneberg claimed that left hemisphere dominance for language first appears around 2 years of age, after which it grows in strength until adult-like lateralization is reached during adolescence. His model was based on a corpus of data that correlated fairly gross disturbances in acquisition of vocal language with the side of early unilateral brain damage (Basser, 1962), and was widely accepted until quite recently.

However, studies of childhood cerebral asymmetries conducted since Lenneberg's model was proposed have been equivocal regarding the earliest age at which hemispheric asymmetries are found. Some researchers claim to have found adult-like lateralization patterns on simple tasks by 2-1/2 years, while others claim there is a *lack* of obvious lateralization on other tasks until 10-12 years (e.g., Bever, 1971; Bryden & Allard, 1977; Kimura, 1963; Knox & Kimura, 1970; Satz, Bakker, Teunissen, Goebel, & Van der Vlugt, 1975; Hiscock, Note 2). Findings have also been equivocal regarding ontogenetic changes in strength of lateralization (Porter & Berlin, 1975). On the one hand, there are suggestions of developmental increases in lateralization elicited by certain tasks; on the other, there are reports of failure to find reliable developmental changes in strength of lateralization (e.g., Berlin, C. I., Hughes, Lowe-Bell, & Berlin, H. L., 1973; Bryden & Allard, 1977; Inglis & Sykes, 1967; Mirabile, Porter, Hughes, & Berlin, C. I., 1978; Bryden, Allard, & Scarpino, Note 3).

More important, infants' perceptual-cognitive abilities, which must serve as the foundation for the more complex lateralized functions found in older children and adults, do appear to be lateralized very early in life. The clinical developmental literature since the 1960's shows that even in infants or very young children, left more often than right hemisphere damage delays or disturbs finer aspects of language development than were assessed in Basser's (1962) original study (e.g., Aicardi, Amsili, & Chevrie, 1969; Alajouanine & Lhermitte, 1965; Annett, 1973; Byers & McLean, 1962; Hécaen, 1976; Kinsbourne, 1975; see review by Entus, Note 4). In addition, individuals who suffered unilateral perinatal cortical damage show subtle lateralized deficits in functioning as adults, even though by adulthood they had adequately developed the *general* class of behaviors usually associated with the hemisphere that had been damaged. That is, right infantile hemidecorticates display subtle but reliable visuo-spatial deficits, and left infantile hemidecorticates show subtle syntactic language deficits (e.g., Dennis & Kohn, 1975; Dennis & Whitaker, 1976; Kohn & Dennis, 1974; McFie, 1961; McFie & Thompson, 1971). Further support for very early functional lateralization is found in the fact that children with less-than-complete early left hemisphere damage often develop or retain language functions in the undamaged speech areas of the left hemisphere, rather than developing right hemisphere language (Milner, 1974; Rasmussen & Milner, 1977).

In order to uncover the developmental basis and time-course for lateralized behavioral functions, however, it would be essential to have information about *normal* lateralized development in addition to information about the clinical effects of unilateral neurological damage. Neuroanatomical and behavioral studies of infants provide support for early cerebral lateralization in neurologically-normal cases. Very early in development, there is evidence of important physical asymmetries in certain language-relevant areas of the neocortex. In young infants, the temporal and frontal lobe language areas in the left cerebral hemisphere are larger than the analogous areas in the right cerebral hemisphere (Witelson & Pallie, 1973), as is the case for adults (Geschwind & Levitsky, 1968), even by 29 gestational weeks (Wada, Clark, & Hamm, 1975). In addition, behavioral data suggest an early lateral asymmetry for manual control, which is related to familial handedness factors. Infants from strongly right-handed families show a right hand advantage for grasp duration by at least 2 months of age when small objects are placed in both hands simultaneously, and a right hand advantage for reaching behavior by 5 months when objects are placed at midline or on both sides of midline (Hawn, Note 5; Hawn & Harris, Note 6).

Functional hemispheric asymmetries in auditory perception have also recently been found in infants. Greater amplitude auditory evoked responses are found over the left than the right hemisphere when speech stimuli are presented, and greater amplitude responses are found over the right than over the left hemisphere when acoustically complex nonspeech auditory stimuli are presented to infants as young as a few days old (Molfese, D. L., Freeman, & Palermo, 1975; Molfese, D. L., Nunez, Seibert, & Ramaniah, 1976; Molfese, D. L., Note 7; Molfese, D. L., & Molfese, V. J., Note 8), which generally parallels the adult pattern of neurocortical response asymmetries (e.g., McAdam & Whitaker, 1971; McKee, Humphrey, & McAdam, 1973; Molfese, D. L., Freeman, & Palermo, 1975). Also, 2-1/2-month-olds show the adult pattern of ear asymmetries in detecting phonetic contrasts among speech stimuli and among various musical instruments playing a single note, when they listen to continuously-presented dichotic stimuli in a high-amplitude-sucking-rate (HAS) habituation/dishabituation paradigm (Entus, 1977; Note 4).

By at least 3 months, infants are also apparently lateralized for making intraclass discriminations of auditory stimuli based on short term storage, or perceptual persistence, of stimulus characteristics. Glanville and Best (Glanville, Best, & Levenson, 1977; Best & Glanville, Note 9) found the adult pattern of ear asymmetries in 3-month-olds' ability to discriminate among phonemes and among musical notes, in a heart rate orienting (OR) habituation/dishabituation study that used discrete stimulus presentations separated by fairly long ($\underline{M}$ = 20 sec) interstimulus intervals. This OR study provides evidence of lateralized differences in 3-month-olds' auditory short term memory, since the design required subjects to retain the relevant stimulus characteristics in order to make phoneme and timbre discriminations.

Thus, the human neocortical hemispheres are, to some extent, functionally and anatomically asymmetrical from very early in life. Yet we still do not know in any sense exactly what processes are lateralized in the infant brain, let alone how that lateralization comes about and develops. In order to better understand the what, and to gain insight on the how, of the ontogenetic origin of auditory perceptual lateralization, it is important to know what (if any) qualitative changes in the behavioral manifestations of functional asymmetry occur during early infancy.

To date, there has been little systematic developmental research on infant cerebral asymmetries. In the earliest studies of infant auditory lateralization, both D. L. Molfese and Entus looked at a fairly wide age range in early infancy, but their age difference analyses were based on small numbers of infants at given ages, and on statistical tests that are relatively insensitive to age differences (e.g., median-split t-tests). A step toward tracing the early ontogeny of auditory lateralization has been made by D. L. Molfese (Note 7), who found some interesting differences between 2- and 5-month-olds in patterns of asymmetrical auditory evoked responses to voicing and place of articulation differences among stop consonants. Deeper insight into the nature of the early development of lateralized perceptual/cognitive behavior, particularly in linguistically-relevant versus nonlinguistic perception of auditory stimuli, would follow from more overtly behavioral data. The study of early infancy age changes in ear asymmetries for speech and music discriminations under short term memory load offers the possibility of gaining richer insight on the nature of lateralization in infant auditory perception. Therefore, in the present investigation of infant dichotic listening abilities, groups of 2-, 3-, and 4-month-olds were tested, using the Glanville et al. (1977) dichotic cardiac OR habituation/dishabituation procedure.

We expected that by at least 3 months infants would show a right ear advantage (REA) in discriminating among speech phonemes and a left ear advantage (LEA) in discriminating among instruments playing a single note. This would reflect the adult pattern of perceptual ear asymmetries, and would corroborate earlier behavioral evidence of infant auditory asymmetries. However, predictions about the lateralized perceptual behavior of 2-month-olds are more difficult. Although D. L. Molfese's work suggests that 2-month-olds should show the same pattern as adults, other recent findings suggest that they may differ from older infants in ear asymmetries, at least for more overt behavioral measures of consonant discrimination. For example, in an attempted replication of Entus' (Note 4) study, Vargha-Khadem and Corballis (Note 10) failed to obtain an REA for speech discriminations in a group of 2-month-olds.

It may very well be that infants under about 3 months fail to use adequate coding and storage skills (or fail to extend perception temporally) to make consonant discriminations under conditions of long ISIs. It has recently been shown that infants younger than 3-4 months fail to make certain phonetic discriminations under conditions of short term memory load, even though they make the analogous discriminations under minimal memory load (Morse, 1978). If infants under 3 months do show a deficit in phonetic coding and/or storage, we would not expect an REA in 2-month-olds on a memory-load task. In fact, if 2-month-olds are generally immature with respect to those behavioral qualities, we would not expect to find any evidence of consonant discrimination by either hemisphere in that age group. No empirically- or theoretically-motivated predictions could be made about the ontogeny of the LEA for music timbre discriminations because, with the exception of the Entus (1977; Entus, Note 4) and Glanville et al. (1977; Best & Glanville, Note 9) lateralization studies, there are no data on young infants' perception of musical timbre. Moreover, no theoretical perspective has directly addressed issues about the nature of very early timbre perception.

METHOD

Subjects

A total of forty-eight 2-, 3-, and 4-month-old infants (eight infants in each Age X Sex subgroup) completed the study. Mean age for the three groups was, respectively: 62.47 days (S.D. = 3.37, range = 55-66), 92.69 days (S.D. = 4.30, range = 87-103), and 123.5 days (S.D. = 4.31, range = 115-130). The subjects were recruited via mailings to recent parents whose names were listed in local newspaper birth announcements. Subjects were screened for birth complications, and none were on medication at the time of testing.[2]

In all, 132 infants participated in the study; the overall attrition rate was thus 63.64%.[3] Crying was the most frequent reason for an infant's failure to complete the experimental session (32 Ss). Other causes included sleeping (12 Ss), interference of excessive squirming with heart rate recording (11 Ss), parental interference (4 Ss), experimental error (including infants outside the appropriate age range -- 20 Ss), equipment failure (4 Ss), or infant illness (2 Ss).

Procedure

The 48 participant infants completed four dichotic OR habituation/dishabituation tests: a right-ear and a left-ear test for phonetic discrimination, and a right-ear and left-ear test for musical timbre discrimination, to test for ear differences within each stimulus type. During all tests the infants' cardiac rate responses were monitored on a Grass model 7 polygraph via three Beckman biopotential electrodes. Two recording electrodes were taped to the infant's chest, one inch (2.54 cm) above each nipple, and one ground lead was taped to the infant's left earlobe. Raw electrocardiograms (ECG's) were collected on one polygraph channel through a Grass model 7P122 preamplifier, while heart rate in beats-per-minute (BPMs) was simultaneously recorded on a second channel through a Grass model 7P4 tachograph preamplifier. The heart rate response of interest was an orienting response (OR), which is a stimulus-elicited phasic deceleration.

Each habituation/dishabituation test consisted of 10 trials. On each trial an auditory dichotic stimulus was presented once at a comfortable but obviously audible 68 db (scale C, Brüel-Kjaer sound level meter, model 2203) over lightweight Sennheiser HD400 open-air stereo headphones. Intertrial intervals were varied randomly from 15 to 25 seconds (M = 20 sec), to avoid eliciting temporal conditioning. For each of the four tests, a dichotic stimulus pair was named as the habituation pair, and was chosen from the one stimulus set (speech syllables or music notes) that was to be tested on that test. During each of the first nine trials of each test, the dichotic habituation stimulus pair for that test was presented once, such that the right ear received one member of the pair while the left ear received the other member simultaneously. The tenth trial of each test was a test trial, in which one ear again received its habituation stimulus while the other ear received a novel test stimulus taken from the same stimulus set as the habituation pair. Pilot testing as well as earlier experimental work (Glanville, Best, & Levenson, 1977; Best & Glanville, Note 9) indicated that the cardiac OR would habituate during the first nine trials, or habituation portion, of the tests. Recovery of the OR on the test trials (stimulus change) would then indicate dishabituation, and therefore would support the interpretation that the infants had detected the stimulus change.

All infants received two tests for each of the two stimulus types. For one of the two tests within a stimulus type, the novel test stimulus on trial 10 was presented to the left ear; for the other test within a stimulus type the novel stimulus on trial 10 was presented to the right ear.[4] The speech syllables were computer-synthesized three-formant one-syllable tokens of each of the six English stop consonants followed by the vowel /a/, all of the syllables being highly identifiable to adults, based on prior testing. The syllables were each 350 msec in duration, and had initial 45 msec formant transitions, without stop burst cues, which distinguished place of articulation. Speech syllable set A consisted of /ba/ and /da/ as the dichotic habituation pair, and /ga/ as the novel test stimulus. Studies of speech perception in young (1- to 4-month) infants show they can discriminate among syllables containing these voiced stop consonants in the context of various vowels, according to the adult phoneme categories for place of articulation (e.g., Eimas, 1974a, 1974b, 1975; Miller & Morse, 1976; Moffitt, 1971; Morse, 1972, 1974). Speech syllable set B consisted of /pa/ and /ta/ as the dichotic habituation pair, with /ka/ as the novel test stimulus. Music note stimuli were 600 msec (75 msec rise and fall times) Minimoog-synthesized renditions of the note A above middle C (440 Hz) by various musical instruments. Music note stimulus set A was piano and brass stimuli as the habituation pair, with reed as the novel test stimulus; set B had organ and string as the habituation pair, and flute as the novel test stimulus.

Presentation orders of the two stimulus types and of the two stimulus sets within each stimulus type (A vs. B) were counterbalanced between subjects. The order in which the two ears received the novel stimulus on trial 10 for the four tests was counterbalanced within subjects. There was a one-minute pause between tests within each stimulus type to reverse the headphone channels, and a five-minute pause between test blocks for the two stimulus types.

Data reduction

Mean heart rate in BPM was calculated on each trial for the 5 second prestimulus period (preceding stimulus onset) and the 5 second poststimulus period (following stimulus offset) for individual infants. Each of these 5-second means was determined by finding the average BPM for each of the 5 seconds in the given period, and then averaging those five means. Five seconds was chosen as the measurement interval because many studies of the second-by-second course of poststimulus cardiac deceleration in young infants (1-1/2- to 4-month) indicate that peak deceleration is typically achieved around the fifth poststimulus second (Graham & Jackson, 1970; Berg, K. M., Berg, W. K., & Graham, Note 11; Hatton, Note 12). Analyses based on the data for both the average heart rate during the prestimulus period and the average rate during the poststimulus period are hereafter referred to as the Period analyses. The Difference Score analyses were based on the average change from prestimulus to poststimulus BPM on each trial. Heart rate Difference Scores for individual infants were calculated on each trial by subtracting the poststimulus mean BPM from the prestimulus mean, so that in those Difference Score analyses a positive difference score reflects a cardiac deceleration (OR) following stimulus presentation.

RESULTS

Overall group analyses

Analyses of the habituation trials for all tests

To determine whether habituation of the cardiac OR occurred in the first nine trials of all tests, two overall within-subjects analyses of variance on those trials were performed for the total sample. A Stimulus Type (speech syllables or music notes) X Ear (left or right ear to be tested on trial 10) X Trial (1 through 9) X Gender X Age (2- vs. 3- vs. 4-month-olds) ANOVA (Analysis of Variance) was carried out on the heart rate Difference Scores, as well as a Stimulus Type X Ear X Trial X Gender X Age X Period (prestimulus vs. poststimulus mean BPM) ANOVA on the Period data.

The habituation trials analyses indicated that the cardiac OR habituated in all four tests. The Difference Score analysis revealed a significant Trial effect, $\underline{F}_{(8,336)} = 2.27$, $\underline{p} < .025$, supporting a reliable decrease over trials in the size of the cardiac OR (see Figure 1). The significant Period effect in the Periods analysis, $\underline{F}_{(1,42)} = 91.37$, $\underline{p} < .0001$, signifies that the infants showed a general OR to the habituation stimuli. Furthermore, the significant Trial X Period interaction, $\underline{F}_{(16,336)} = 3.18$, $\underline{p} < .0001$, suggests that prestimulus heart rate remained relatively constant over trials during habituation, but that the magnitude of poststimulus heart rate change from prestimulus levels diminished over trials (see Figure 2). Simple effects tests were run to break down this and all subsequent interactions, since correct statistical interpretation of interactions requires knowledge of those sources of variance which contributed to it significantly (Kirk, 1968). The results of the simple effects tests for the Trial X Period interaction showed that while the prestimulus heart rate differed from the poststimulus rate on trial 1, $\underline{F}_{(1,378)} = 9.31$, $\underline{p} < .005$, there was no significant pre- versus post-stimulus difference on trial 9. This finding indicates there was an OR on trial 1, but not on trial 9, which supports the interpretation that the cardiac OR habituated during the first nine trials of the four tests. It is of some interest that the overall habituation trials analyses failed to provide evidence of reliable Stimulus Type differences in the shape of cardiac OR habituation on this task.[5] The overall analyses of the habituation trials provide no evidence, moreover, of Age differences in the shape of cardiac OR habituation.[6]

Dichotic test trial analyses

The results of the statistical tests most relevant to the questions about infant cerebral asymmetries for speech and timbre discriminations are the Difference Score and Period analyses of the last habituation trial (9) versus the test trial (10), and the analyses on the test trial alone, for all four tests. Results from these analyses supported the predictions of an REA for speech phoneme discriminations and an LEA for musical timbre discriminations by the 2-, 3-, and 4-month-old infants tested. Within-subjects ANOVAs for Stimulus Type X Ear X Trial X Gender X Age were performed on the Difference Score data (and X Period for the Period data) for trials 9 versus 10, and for Stimulus Type X Ear X Gender X Age data on trial 10. These particular tests were run to assess whether there was recovery of the cardiac OR on trial 10 relative to the cardiac response on trial 9, and to learn whether the trial 10 dishabituation was related to Ear receiving the novel test stimulus as well as to the Stimulus Type.

The significant trial 10 Stimulus Type X Ear interaction for the difference score analysis, $\underline{F}_{(1,42)} = 10.47$, $\underline{p} < .003$, is illustrated on the right-hand side of Figure 1, as is the significant trial 9 versus 10 Stimulus Type X Ear X Trial interaction, $\underline{F}_{(1,42)} = 4.25$, $\underline{p} < .05$. Both findings suggest a speech REA and a musical timbre LEA in the pattern of trial 10 cardiac ORs. The predicted pattern of dishabituations was further suggested by a significant Stimulus Type X Ear X Trial X Period interaction in the Periods analysis of trials 9 versus 10, $\underline{F}_{(1,42)} = 4.25$, $\underline{p} < .05$.

A series of simple effects tests was performed on the test trial results just reported, to determine which sources of variance produced the interactions. The cardiac OR habituated equally by trial 9 for all four tests, since no reliable cardiac decelerations were found on trial 9 for any test. However, there was a significant ear by stimulus type difference in the size of the cardiac OR on trial 10, favoring greater dishabituation for the left than for the right ear tests of music discrimination, $\underline{F}_{(1,84)} = 4.98$, $\underline{p} < .05$, and greater dishabituation for the right than for the left ear tests of phonetic discrimination, $\underline{F}_{(1,84)} = 5.31$, $\underline{p} < .025$. Further simple effects tests of the Difference Score data for the test trial alone revealed that, indeed, the right ear (left hemisphere) stimulus change produced a larger OR for the speech tests than the music timbre tests, $\underline{F}_{(1,84)} = 7.88$, $\underline{p} < .01$, and the left ear (right hemisphere) stimulus change produced a larger OR for music than for speech tests, $\underline{F}_{(1,84)} = 4.11$, $\underline{p} < .05$. Moreover, simple effects tests of the trials 9 versus 10 Difference Score data found significant dishabituation of the cardiac OR only for the right ear speech tests, $\underline{F}_{(1,84)} = 8.69$, $\underline{p} < .005$, and the left ear music timbre tests, $\underline{F}_{(1,84)} = 6.71$, $\underline{p} < .025$.

Simple effects tests of the trials 9 versus 10 interactions revealed that the trial 9 prestimulus versus poststimulus heart rate differences were not significant (no reliable cardiac ORs) for any of the four tests, nor were there any significant differences among the trial 9 responses on the four tests. So it appears that the cardiac OR had habituated by trial 9, and to an equal extent for all tests. On trial 10 the poststimulus rate was lower than the prestimulus rate (indicating dishabituation of the cardiac response) only for the right ear speech test, $\underline{F}_{(1,84)} = 8.69$, $\underline{p} < .005$, and the left ear music test, $\underline{F}_{(1,84)} = 6.71$, $\underline{p} < .025$. Therefore, the trial 10 analyses, and even more strongly the trials 9 versus 10 analyses, very strongly support the prediction that a group of young infants between 2 and 4 months show the adult pattern of cerebral asymmetries in use of auditory short term memory for making intraclass stimulus discriminations. That is, these infants showed a left hemisphere advantage in making speech sound discriminations, and a right

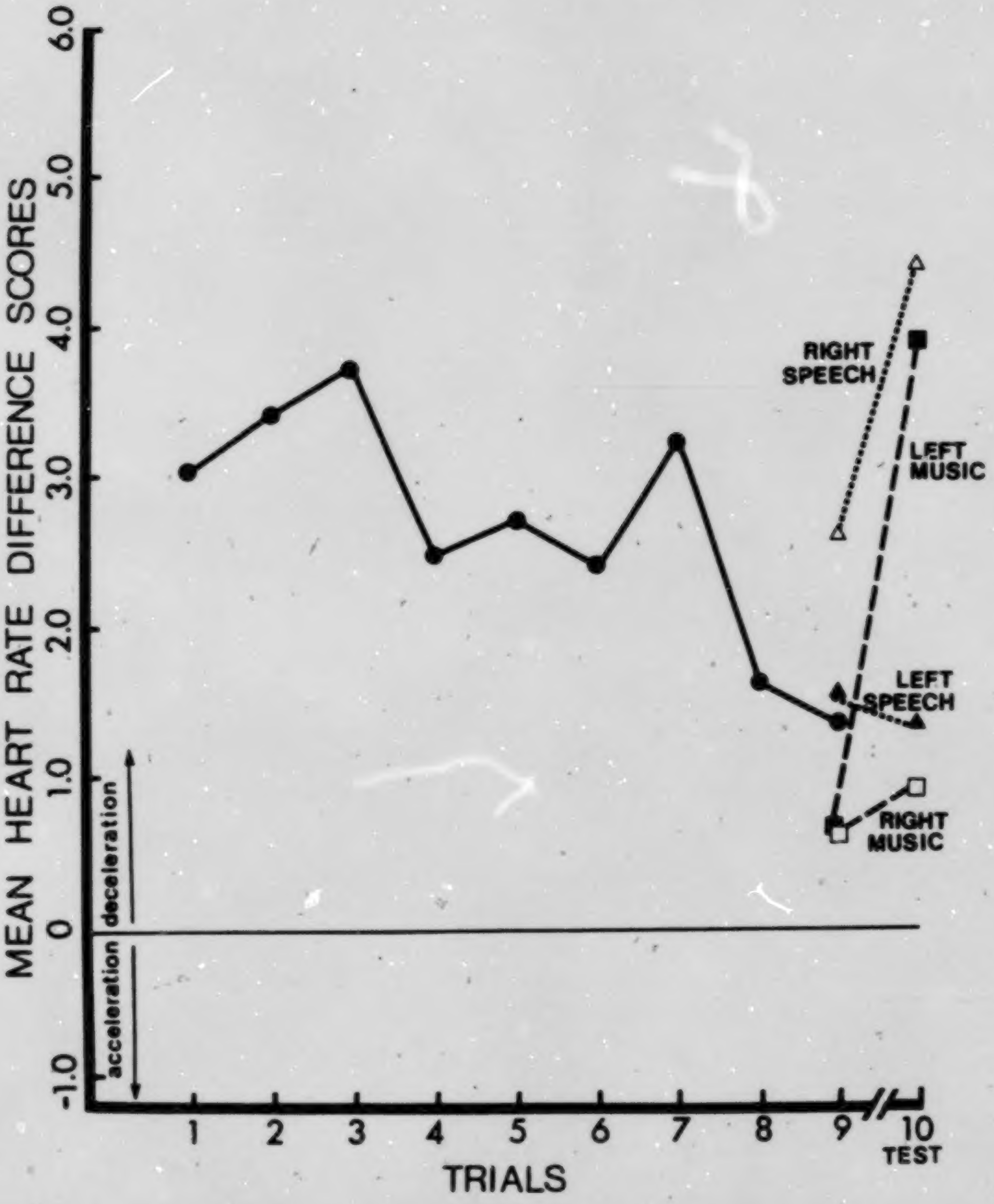


Figure 1. Habituation of the cardiac OR over the habituation trials for all conditions, and recovery of the OR on the test trial for each condition.

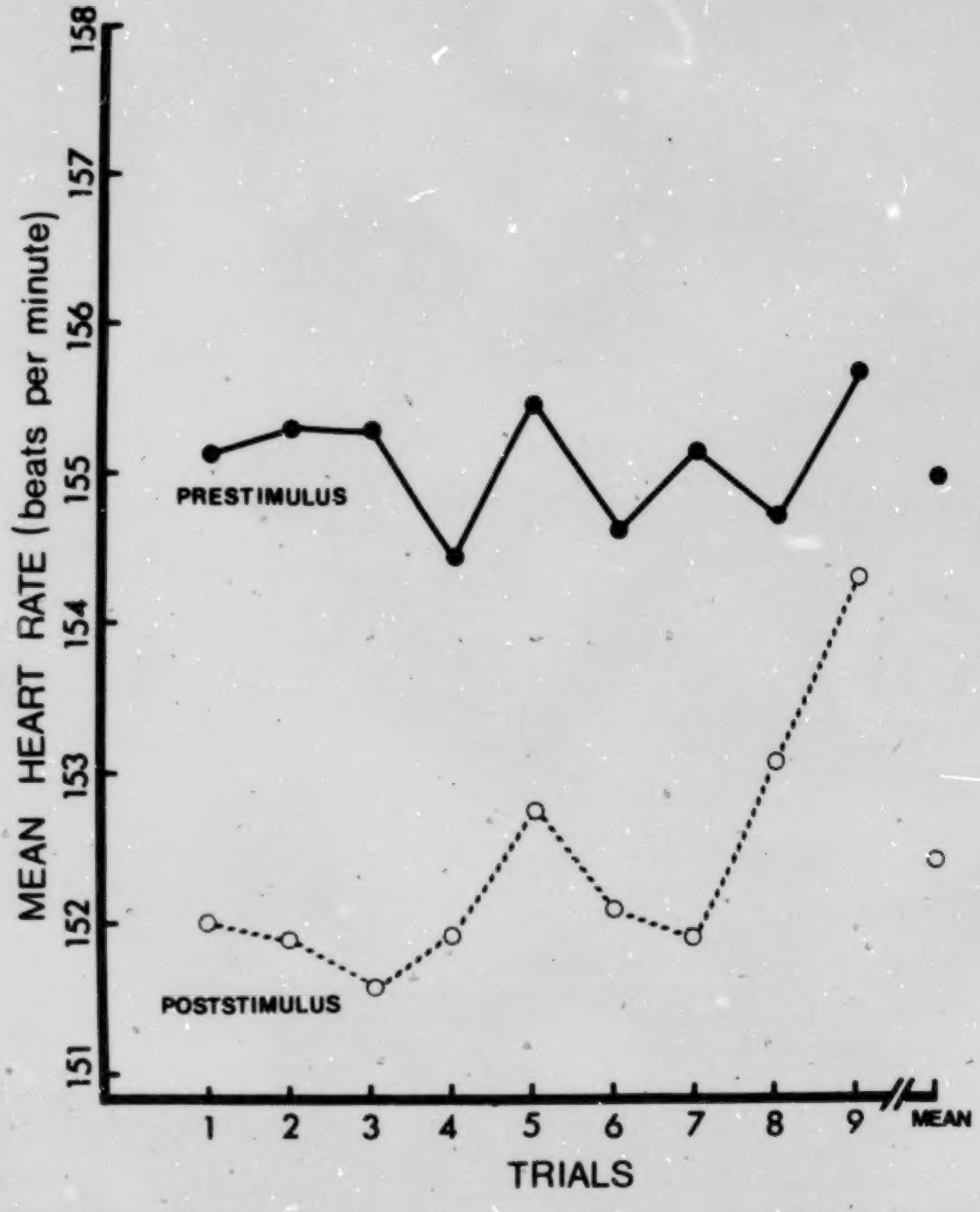


Figure 2. Mean heart rate in BPM for the 5 sec prestimulus Period and the 5 sec poststimulus Period on each trial, and the grand mean for all habituation trials.

hemisphere advantage in making music timbre discriminations.

Within-age-group analyses

The lack of significant Age effects in the overall ANOVAs for trials 9 versus 10 suggests that the overall pattern of cardiac dishabituation results is similar among the three age groups. However, it has been shown that ANOVAs involving an age factor with more than two levels (or in fact *any* factor with more than two levels) may sometimes fail to yield any significant effects for that factor even when there are reliable differences among the levels of that factor, because of the way the error term is derived in traditional ANOVA formulae (Hale, 1977). Since we wanted to know whether the general pattern of ear asymmetries in test trial dishabituations held up to the same extent in each age group, and felt that the overall ANOVAs may have missed some possible reliable age differences, we ran further ANOVAs separately for each age group on the Difference Score and Period data for trials 9 versus 10, and for trial 10 alone.

Two-month-olds

The dishabituation results for the 2-month-olds suggests that these subjects have right hemisphere superiority for making musical timbre discriminations. However, there is no reliable evidence that this age group makes phonetic discriminations with *either* hemisphere under the short term memory load of the paradigm used. Although this age group showed evidence of cardiac habituation by trial 9 of all four tests, in that there was no significant cardiac OR on trial 9 for any test, the trial 10 Difference Score analyses indicate cardiac OR dishabituation *only* for the left ear musical timbre change. The significant Stimulus Type X Ear interaction, $F_{(1,14)} = 4.19$, $p < .05$, shown in Figure 3, supports the claim just made. Simple effects tests of the trial 10 results show a Stimulus Type difference in the left ear response on trial 10, with a larger cardiac deceleration for music than for speech, $F_{(1,28)} = 3.92$, $p < .056$. There was not a significant Stimulus Type difference in magnitude of cardiac OR for the right ear on trial 10. The trial 10 Period analysis yielded further support for the music LEA in the 2-month-olds, via a significant Stimulus Type X Ear X Period interaction, $F_{(1,14)} = 4.19$, $p < .05$. Again, simple effects tests revealed a significant poststimulus cardiac deceleration to the test stimulus *only* for the left ear music test, $F_{(1,28)} = 4.21$, $p < .05$. The trial 10 Period differences for the other three test conditions fell far short of significance, failing to support significant recovery of the cardiac OR to any of those stimulus changes.

The pattern of test trial results among the 2-month-olds provides behavioral evidence of right hemisphere specialization for music timbre discriminations under short term memory load. There is no evidence of left hemisphere specialization in this age group for speech discriminations under memory load, nor in fact is there evidence of *any* phoneme discrimination under the memory constraints of the experimental task used. Other reports that infants under three or four months of age fail to show evidence of phoneme discriminations under conditions that place a load on short term memory (cf. Morse, 1978) corroborate the speech discrimination data in our youngest group.

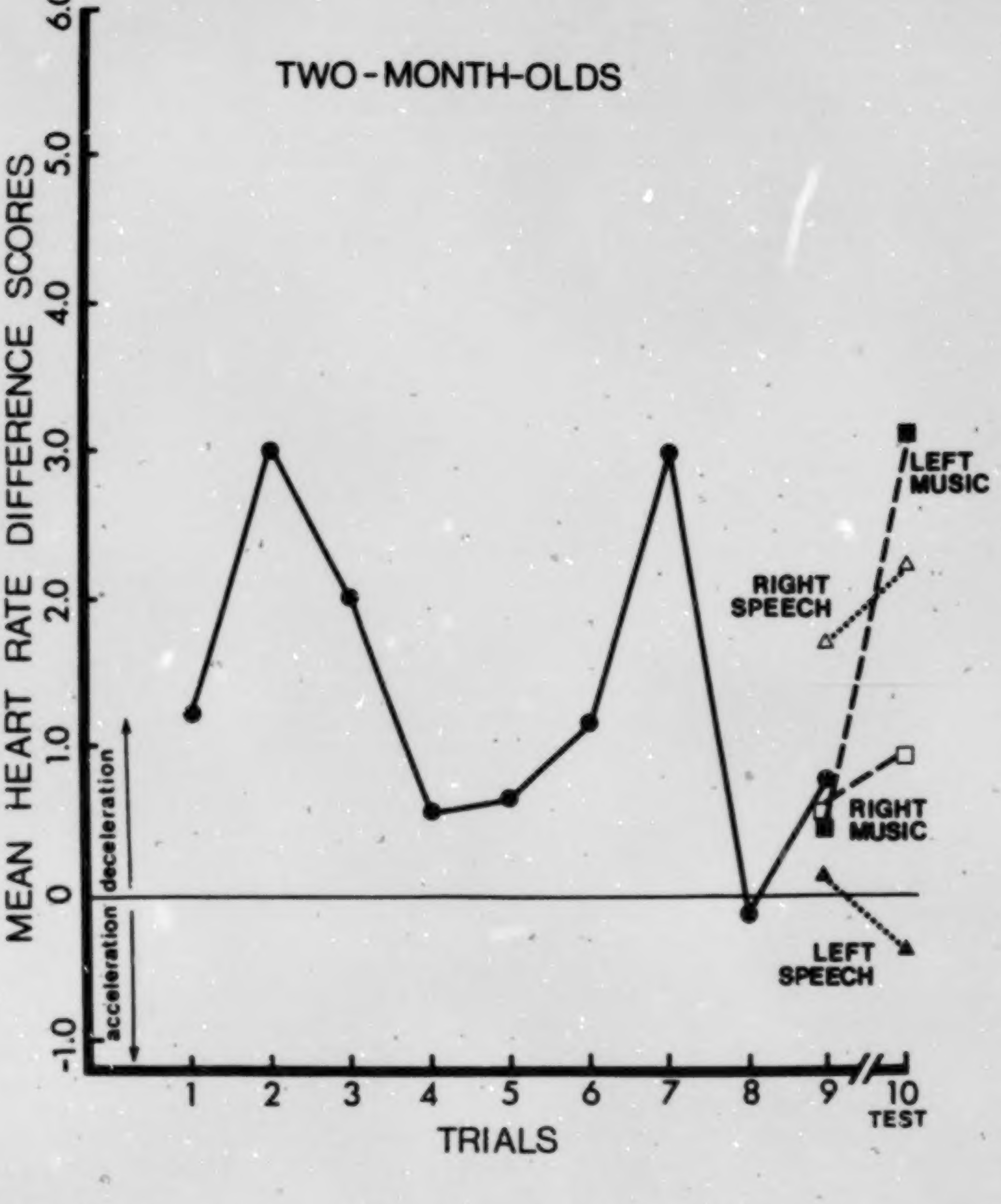


Figure 3. OR habituation for all conditions, and OR recovery for the test trial in each condition, in the 2-month-olds.

Three-month-olds

The results of the test trial analyses for the 3-month-olds replicate the pattern of ear asymmetries found by Glanville et al. (1977; Best & Glanville, Note 9). The findings for this age group are shown in Figure 4. This age group showed habituation of the cardiac OR by trial 9, in that there were no reliable cardiac decelerations on trial 9 for any of the four tests. Furthermore, the test trial analyses provide evidence of an REA for phonetic discriminations, and an LEA for musical timbre discriminations, in this age group. Simple effects tests of the Stimulus Type X Ear X Trial interaction from the results of the Difference Score ANOVA for trials 9 versus 10 support a significant cardiac OR dishabituation by the 3-month-olds on trial 10 for the right ear speech test, $F_{(1,28)} = 8.36$, $p < .01$, and the left ear music test, $F_{(1,28)} = 11.19$, $p < .005$. There was no significant cardiac OR dishabituation by this age group on trial 10 for the right ear music test or the left ear speech test.

Four-month-olds

As with the 3-month-olds, the analyses indicated an REA in this age group for speech discriminations and an LEA for musical timbre discriminations. Lack of significant cardiac deceleration on trial 9 of all tests with the 4-month-olds suggests that habituation occurred by that trial for all four tests. Of greater interest regarding the experimental predictions, the speech REA and music LEA are supported by the significant Stimulus Type X Ear X Period interaction in the Periods analysis for trial 10 cardiac responses, $F_{(1,14)} = 4.74$, $p < .05$ (see Figure 5). Simple effects tests support the conclusion that the cardiac OR dishabituated on trial 10 only for the left ear music test, $F_{(1,28)} = 5.73$, $p < .025$, and the right ear speech test, $F_{(1,28)} = 12.05$, $p < .001$. The trial 10 results for the other two test conditions yielded no evidence of cardiac OR dishabituation. Thus the 4-month-olds, like the 3-month-olds, show the adult pattern of ear asymmetries for making phoneme and timbre discriminations under short term memory load.

DISCUSSION

The major results of this study replicated those of an earlier dichotic habituation/dishabituation investigation with 3-month-olds (Glanville, Best, & Levenson, 1977; Best & Glanville, Note 9). Within the total sample of 2-, 3-, and 4-month-old infants in the present experiment, there is substantial evidence in the analyses of data for the total group for a left hemisphere advantage in discriminating phoneme contrasts between speech syllables, and for a right hemisphere advantage in discriminating timbre contrasts between music notes. Given the short term memory requirements of the paradigm used, these findings reflect the adult pattern of cerebral asymmetries in young infants for the short term persistence of the perceptual qualities essential to discrimination of the speech and music contrasts tested. The general findings suggest that even young infants somehow dichotomize their auditory world according to the two basic "natural" categories adults use. These young infants perceived the two types of complex acoustic patterns they heard in this test in qualitatively different manners, convergent with adult categori-

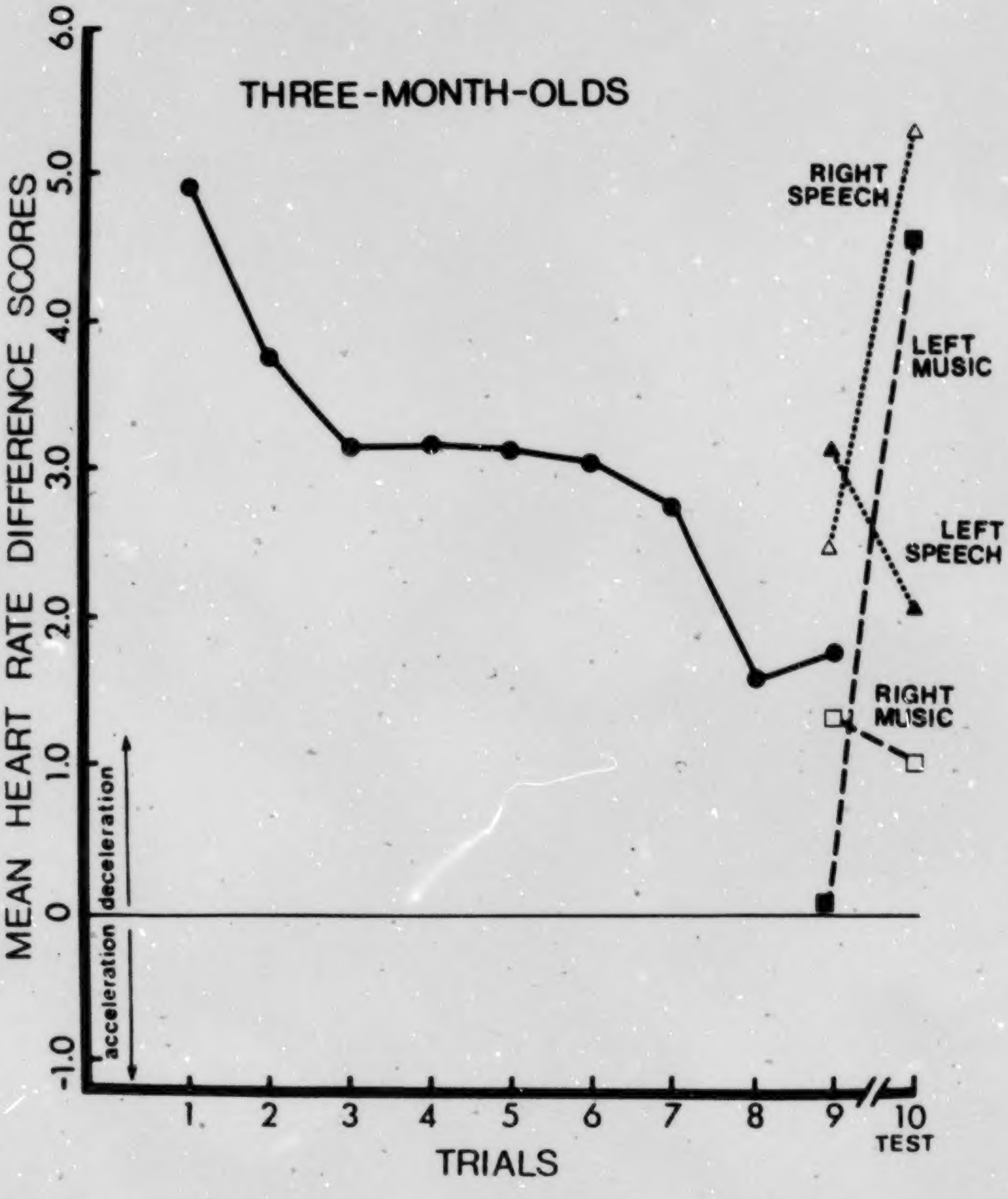


Figure 4. OR habituation for all conditions, and OR recovery for the test trial in each condition, in the 3-month-olds.

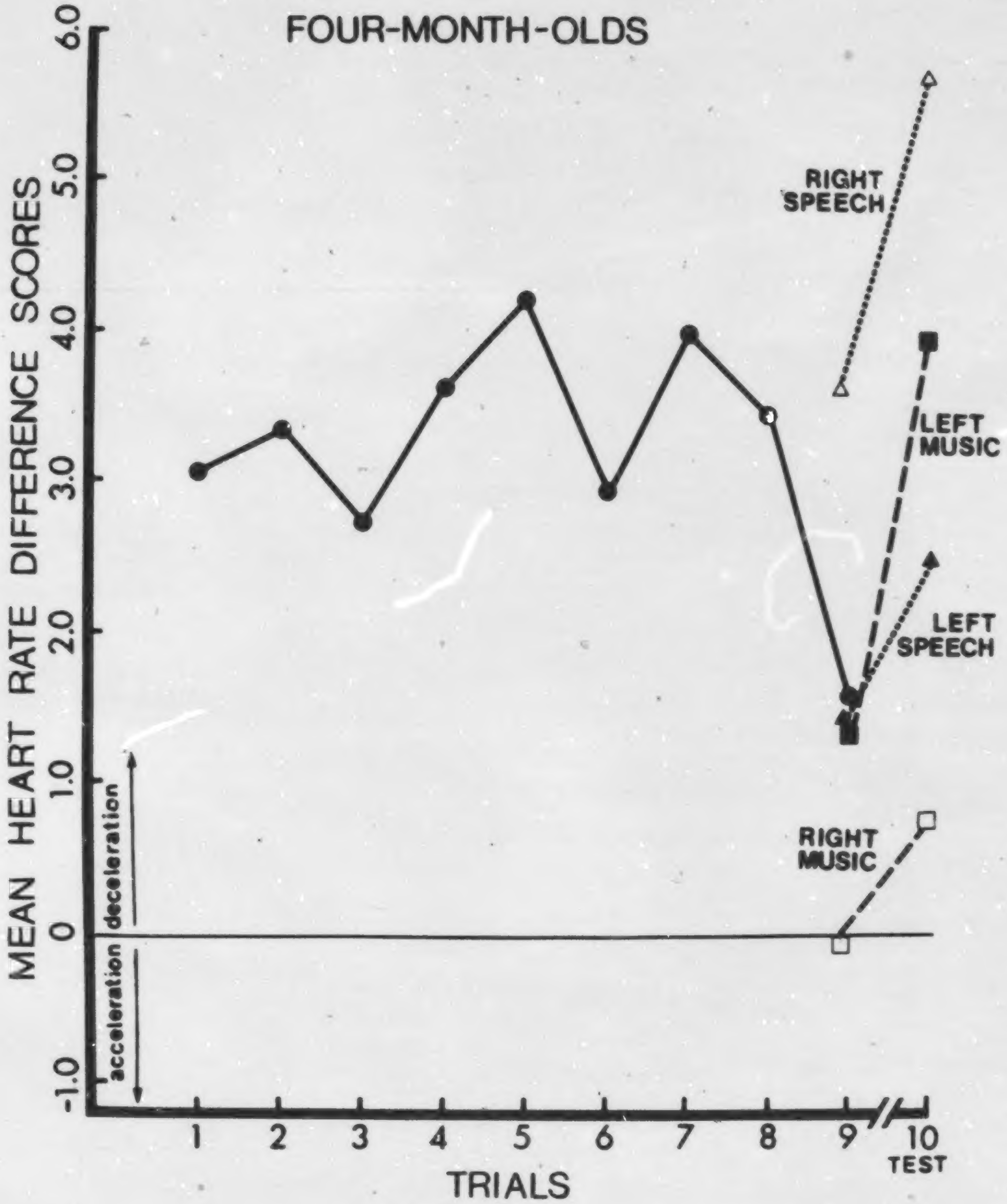


Figure 5. OR habituation for all conditions, and OR recovery for the test trial in each condition, in the 4-month-olds.

zations of the stimuli as human speech sounds or as complex nonspeech sounds (musical timbre).

Two basic stimulus categories in infant auditory perception

The present findings bring up the important and difficult question of how it is that very young infants with their limited auditory experience, and lack of "meaningful" semantic and syntactic language abilities (at least, so far as is known), differentiate between speech and music notes in their neurophysiologically-based perceptual behavior. The two classes of stimuli are quite similar in many respects. Both stimulus types have acoustic complexity, and they share many features which have been found quite effective in gaining infants' attention. For example, acoustic descriptors of both speech and music include a broad range of frequencies below 8 KHz, harmonic or periodic structure, some formant or bandpass-filtered structure, moderate rise time, and onset transients in frequency and amplitude. Moreover, the perceptual qualities associated with adults' phonetic and music discriminations are based on stimulus differences that involve a complex of contrasts along several acoustic dimensions, rather than relating to simple contrasts on single acoustic dimensions (e.g., Dorman, Studdert-Kennedy, & Raphael, 1978; Fitch, Halwes, Erickson, & Liberman, 1979; Grey & Gordon, 1978; Best, Morrongiello, & Robson, Note 13; Ehresman, Note 14; Wessel, Note 15).

It is still controversial whether the physical acoustic properties of speech and musical timbre are the cause either of their hemispheric processing differences, or of their basic perceptual/cognitive differences. However, several findings are consistent with a suggestion that acoustic properties play at least some role, and perhaps an important one, in cerebral asymmetries for perception of ecologically-relevant auditory stimuli such as speech. First, it has been found that under many conditions there is an apparent lack of cerebral asymmetry in adults for perception and identification of vowels (Shankweiler & Studdert-Kennedy, 1967; Studdert-Kennedy & Shankweiler, 1970). This finding would seem surprising in that vowels are crucial components of human languages and are important in perception of ongoing speech. By that token, if the basis for cerebral lateralization were organized along "linguistic/nonlinguistic" lines, we would have expected to find an REA for vowels, as is found for perception of consonants. However, comparison of the acoustic structure of isolated vowels versus the acoustic properties thought to be associated with consonants suggests that acoustics may play some part in the laterality differences for the two phoneme classes. Identification of vowels is at least partially dependent on the frequency relationships among the three low-frequency formants, an acoustic property that is similar in some ways to, and different in others from, the acoustic properties of both consonants, which usually yield an REA, and of music notes and chords, which often yield an LEA. Conversely, some clearly nonlinguistic tasks have been found to produce an REA rather than the LEA that would be expected if lateralization were based on a "linguistic/nonlinguistic" dichotomy. These findings of a nonlinguistic REA have typically involved recognition or comparison of complex nonspeech sounds that had to be judged according to acoustic characteristics similar in some ways to consonantal characteristics (such as rapid changes in intensity and spectral characteristics), again supporting the proposition that acoustic properties may in part affect ear asymmetries in dichotic tasks (e.g., Cutting, 1974; Halperin, Nachshon, & Carmon, 1973; Cutting, Note 16).

It seems at least plausible that the effect of acoustic properties on adult hemispheric processing asymmetries could be better appreciated in an ontogenetic context. That is, we might gain insight on the nature of adult brain lateralization if we knew more about its origins in infancy. Investigation of infant cerebral asymmetries may be a fruitful avenue by which we can discover more about the ontogeny of perception, in terms of what stimulus information specifies "speech-like" versus "nonspeech" to the infant. This suggestion is supported by the recent finding of an LEA for discrimination of steady-state vowels by 3-1/2-month-olds who showed an REA for discrimination of consonant noise cues (Best, Note 17). The study of infant cerebral asymmetries may help in general to illuminate details in the ontogeny of qualitative perceptual/cognitive differences for the major subcategories of stimuli within the auditory modality.

Early age differences in lateralized auditory behavior

As for possible age differences in auditory cerebral asymmetries in early infancy, the results of the present study indicate that 3- and 4-month-old infants have left hemisphere specialization for making speech stimulus discriminations and right hemisphere specialization for making musical timbre discriminations. The 2-month-olds, however, provided evidence of only right hemisphere specialization for timbre discrimination. As a group, the 2-month-olds did not show evidence of discriminating any phoneme changes under the short term memory constraints of the paradigm used. As a result of this floor effect, they failed to show a left hemisphere advantage for speech.

Research in basic infant speech perception suggests that infants as young as one month can detect a change in the place of articulation (/b/ vs. /d/), or voice onset time (/p/ vs. /b/) of a number of consonant sounds, as long as the syllables are presented continuously at a rate of one or two per second (e.g., Cutting & Eimas, 1975; Eimas, Siqueland, Juszcyk, & Vigorito, 1971; Trehub & Rabinovitch, 1972). Infants this young, however, apparently do not show evidence of making the same phonetic discriminations if the intertrial intervals are lengthened to require short term storage of the stimulus properties. Only by about four months of age do infants show consonant discriminations under conditions of short term memory load (e.g., Miller, Morse, & Dorman, 1977; Miller, Morse, & Dorman, Note 18; Morse, Note 19).

Since the infants in Entus' (1977) dichotic habituation study, at an average age of 2-1/2 months, showed left hemisphere superiority for consonant discriminations under continuous presentation conditions, the findings for 2-month-olds in the present study (as well as the general infant speech perception results) might suggest that infants under 3 to 4 months simply have inadequate memory capacity to make fine-grain intraclass discriminations among acoustically complex sounds under memory constraint. However, the fact that the 2-month-olds made musical timbre discriminations under the same memory constraint suggests that their auditory memory deficit is not a general one, but rather is keyed to speech (particularly consonants). Moreover, the music discrimination was not merely "easier" for them than the phonetic comparison in an absolute sense, since they did show hemispheric specialization for the music tests, favoring the right hemisphere. That is, even for music timbre discriminations, the 2-month-olds show no evidence of making left hemispheric stimulus discriminations under conditions of short term memory load. It appears that the auditory short term memory "deficit" of infants under three months may be specific to consonant discriminations and/or left hemisphere perceptual/cognitive qualities.

The suggestion that 2-month-olds may have a specific difficulty in making phonetic discriminations under short term memory load conditions is not necessarily at odds with findings that infants that young or younger can discriminate phonemes under continuous presentation conditions. The ability to detect a speech stimulus change under continuous presentation conditions may involve mere registration of acoustic contrast, rather than the more abstract phonetic perception found in adults or the persistence of important stimulus qualities in perception or short term memory. When infants two months of age and younger discriminate speech phonemes under continuous presentation conditions they may be responding to simple acoustic contrast at a subcortical or peripheral level, rather than showing asymmetrical neocortical involvement in linguistically-specialized phonetic perception. This explanation would account for Varga-Khadem and Corballis' (Note 10) failure to replicate with 2-month-olds the REA for speech that Entus (Note 4) found with somewhat older infants under continuous presentation conditions. In line with general findings on early age changes for the role of memory in phonetic discrimination, the infants in the Vargha-Khadem and Corballis (Note 10) study did respond to stimulus change trials, suggesting they detected the phonetic contrast, even though they did not show an ear difference in that detection of contrast.

It is important to note here that the discussion about lack of a speech REA in the 2-month-olds is not meant to imply a lack of any speech-related left hemisphere advantage before about three months. The work of D. L. Molfese and his colleagues, for example, indicates lateralization in the degree of evoked neuro-cortical response to the presentation of speech even during the newborn period. Furthermore, even though the present study indicates that the left hemisphere's memory may not reliably support phonetic discriminations under memory load until 3 months, Entus' research (1977; Entus, Note 4) provides evidence of left hemisphere advantage for detection of phonetic contrasts by at least 2-1/2 months under continuous presentation conditions. Perhaps an important change in the nature of the perception of speech occurs between 2-3 months of age, a change from responding to speech as an acoustic stimulus toward perception of speech in a more linguistically-relevant manner. A qualitative change in speech perception would be consistent with other findings of important and quite pervasive perceptual and biobehavioral changes that take place around 2-3 months (Emde & Robinson, 1976).

Individual response patterns:
The dichotic discrete-trials OR habituation test as a measurement tool

An important question in this type of research, for both theoretical and practical concerns, is whether test response patterns can be determined on an individual basis. To assess the feasibility of using this dichotic test to determine individual ear asymmetries, individual infants were classified according to the pattern of their cardiac dishabituations in all four test conditions. Classification of the individual infants was dependent on the amount of their trial 10 dishabituation relative to the trial 9 response for each test condition, and assessed ear differences in amount of cardiac OR recovery between trials 9 and 10 within Stimulus Types. The criteria used to classify an infant as showing discrimination of the test stimulus from the habituation pair were: 1) trial 10 had to show an average deceleration of at least .5 beats-per-minute (BPM), that is, > or = +.5 Difference Score; 2) the trial 10 deceleration had to be at least .5 BPM greater than the trial 9 deceleration during that test sequence, in case any small trial 9 deceleration had occurred; 3) criteria 1) and 2) had to be met for at least one of the two tests within a Stimulus Type for an infant to be classified as discriminating the trial 10 stimulus change in that Stimulus Type; and 4) the ear difference in amount of trial 10 recovery of the cardiac OR relative to trial 9 within a Stimulus Type had to be greater than .5 for an ear difference in discrimination of that Stimulus Type to be recorded.

Of those infants meeting the requirements for trial 10 discrimination for at least one of the two ear tests within a Stimulus Type, all displayed ear asymmetries in OR recovery of at least .8 BPM (most ear asymmetries were much larger in magnitude) within a Stimulus Type, except for one 3-month-old who showed trial 10 discrimination equally for the two ears on the speech tests. The infants who provided evidence of making trial 10 discriminations within Stimulus Types showed the predicted pattern of ear asymmetries. Of the total sample of 48 infants, 33 showed some music timbre discrimination, with a left-ear advantage in 22 of these (2/3). This sample proportion is in line with estimates of the proportion of music LEA in the normal adult population (e.g., Zatorre, 1979). In addition, 33 infants showed some speech discrimination, with a right-ear advantage in 24.[7] Although this proportion is smaller than the 80% of right-handed adults who show a dichotic speech REA (e.g., Kimura, 1967), it is important to remember that among left-handers the incidence of left hemisphere speech dominance is only 60% (e.g., Goodglass & Quadfasel, 1954). The eventual handedness of the infant participants in our study is, of course, unknown. However, the 10% proportion of left-handers in the adult population suggests that perhaps 4-5 of the 48 infant participants will become left-handed. Therefore, the speech REA in 73% of the individual infants is not so far from the expected proportion of speech REA in a representative sample of all adults. The proportion of individuals showing a music LEA versus those showing a music REA, and the proportion showing a speech REA versus those showing a speech LEA or no ear asymmetry, were both significant, $z = 1.92$, $p < .053$, and $z = 2.64$, $p < .01$, respectively. Thus, among the individual infants who discriminated the test trial stimulus changes, there is a strong tendency toward the adult pattern of cerebral asymmetry.

The dichotic cardiac OR habituation/dishabituation paradigm used in this study does appear to be a reliable measure of group cerebral asymmetries for auditory short term memory in infants, at least by three months of age. However, although it is useful for group studies, the individual analyses indicate at present that it may not be sensitive enough for individual assessments. The paradigm might be modifiable for use with younger infants, or for individual testing. Perhaps those goals could be achieved by reducing the duration of the ISIs or by increasing the number of habituation trials, since evidence suggests that the degree of cardiac habituation, and especially cardiac dishabituation, to a stimulus change is correlated with the number of habituation trials (McCall & Melson, 1970).

The paradigm may also be useful in studying the lateralization of auditory functions in other populations of subjects from whom a verbal or motor response cannot be easily obtained, so long as the subjects show cardiac orienting to the stimuli presented. For example, a recent dichotic cardiac habituation/dishabituation study of four classical-autistic preschoolers revealed the same pattern of ear asymmetries found in our infants (Kodera & Best, Note 20). It is extremely difficult to test auditory perception in these children by other, more traditional means because of their asocial behavioral characteristics and failure to spontaneously use language for interpersonal communication. In fact, their peculiar failure to acquire language normally even though their hearing appears to be intact has led several researchers to suggest that autistic children may either have a dysfunctional left hemisphere, or may somehow fail to utilize existing left hemisphere abilities (e.g., Simon, 1975; Takagi, 1972; Tanguay, 1976; Kinsbourne, Note 21; Levy, Note 22). Yet the preliminary findings from the cardiac OR habituation test of autistic children suggest their deficit is *not* a global left hemisphere dysfunction. The recent data on autistics in turn suggest that the discrete-trials cardiac OR procedure may be useful in discovering more about the role of atypical brain lateralization in various perceptual and cognitive developmental abnormalities. In general, the procedure seems a good tool for research on the development of normal infant auditory asymmetries, and may help to shed light on the early ontogeny of perceptual and cognitive brain-behavior relationships.

REFERENCE NOTES

1. Harris, L. J., & Best, C. T. *Sex and hemisphere differences for two performance measures of haptic maze-learning*. Manuscript in preparation, Haskins Laboratories, 1979.
2. Hiscock, M. *Ontogeny of attentional biases in performance*. Paper presented at the Meeting of the American Psychological Association, Washington, D.C., 1976.
3. Bryden, M. P., Allard, F., & Scarpino, F. *The development of language lateralization and speech perception*. Unpublished manuscript, University of Waterloo, 1973.
4. Entus, A. K. *Hemispheric asymmetry in processing of dichotically presented speech and nonspeech sounds by infants*. Paper presented at the Biennial Meeting of the Society for Research in Child Development, Denver, Colorado, 1975.
5. Hawn P. R. *Hand asymmetries in grasp duration and reaching in two- and five-month-old human infants*. Unpublished Masters thesis, Michigan State University, 1978.
6. Hawn, P. R., & Harris, L. J. *Hand asymmetries in grasp duration and reaching in two- and five-month-old human infants*. Paper presented at Meeting of the International Neuropsychology Society, New York, 1979.
7. Molfese, D. L., & Molfese, V. J. *VOT discrimination along phoneme boundaries is not innate*. Paper presented at Meeting of the Midwestern

Psychological Association, Chicago, Illinois, 1978.

8. Molfese, D. L., & Molfese, V. J. *Electrophysiological correlates of categorical perception in infants two to four months old*. Paper presented at the Biennial Meeting of the Society for Research in Child Development, New Orleans, Louisiana, 1977.

9. Best, C. T., & Glanville, B. B. *A cardiac measure of cerebral asymmetries in infants' perception of speech and nonspeech sounds*. Paper presented at the Meeting of the Midwestern Psychological Association, Chicago, Illinois, 1976.

10. Vargha-Khadem, F., & Corballis, M. C. *Cerebral asymmetry in infants*. Paper presented at the Biennial Meeting of the Society for Research in Child Development, New Orleans, Louisiana, 1977.

11. Berg, K. M., Berg, W. K., & Graham, F. K. *Infant heart rate response as a function of stimulus and state*. Unpublished manuscript, University of Wisconsin, Madison, 1968.

12. Hatton, H. M. *Developmental changes in infant heart rate during sleeping and waking states*. Unpublished Doctoral dissertation, University of Wisconsin, Madison, 1969.

13. Best, C. T., Morrongiello, B., & Robson, R. *The perceptual equivalence of the phonetic trading relation between silent gap and F_1 transition extent as cues to the /say/-/stay/ contrast*. Manuscript in preparation, Haskins Laboratories, 1979.

14. Ehresman, D. *A parallelogram model of timbre analysis*. Unpublished Masters thesis, Michigan State University, 1977.

15. Wessel, D. L. Report presented at the Meeting of Mathematical Psychology, San Diego, California, 1973.

16. Cutting, J. E. *There may be nothing peculiar to perceiving in the speech mode*. Paper presented at the Meeting of the Psychonomic Society, St. Louis, Missouri, 1976.

17. Best, C. T. *The role of consonant and vowel acoustic features in infant cerebral asymmetries*. Unpublished Doctoral dissertation, Michigan State University, 1978.

18. Miller, C. L., Morse, P. A., & Dorman, M. F. *Infant speech perception, memory, and the cardiac orienting response*. Paper presented at the Biennial Meeting of the Society for Research in Child Development, Denver, Colorado, 1975.

19. Morse, P. A. *Adaptation, memory, and infant speech perception*. Paper presented at the Meeting of the International Conference on Infant Studies, Providence, Rhode Island, 1978.

20. Kodera, T., & Best, C. T. *Auditory perceptual asymmetries in four classic autistic children parallel normal infant and adult asymmetries*. Unpublished manuscript, Tulane University, 1979.

21. Kinsbourne, M. *The neuropsychology of infantile autism*. Paper presented at the Workshop on the neurobiological basis of autism. National Institute for Neurological and Communicative Disorders and Stroke, Bethesda, Maryland, February, 1976.

22. Levy, J. *Variations in the lateral organization of the brain*. Master Lecture presented at the Meeting of the American Psychological Association. San Francisco, California, 1977.

REFERENCES

Aicardi, J., Amsili, J., & Chevrie, J. J. Acute hemiplegia in infancy and childhood. Developmental Medicine & Child Neurology, 1969, 11, 162-173.

Alajouanine, T., & Lhermitte, F. Acquired aphasia in children. Brain, 1965, 88, 653-662.

Annett, M. Laterality of childhood hemiplegia and the growth of speech and intelligence. Cortex, 1973, 9, 4-33.

Basser, L. S. Hemiplegia of early onset and the faculty of speech, with special reference to the effects of hemispherectomy. Brain, 1962, 85, 427-460.

Benton, A. L. Clinical symptomatology in right and left hemisphere: Discussion. In V. B. Mountcastle (Ed.), Interhemispheric relations and cerebral dominance. Baltimore: Johns Hopkins Press, 1962.

Berlin, C. I., Hughes, L. F., Lowe-Bell, S. S., & Berlin, H. L. Dichotic right ear advantage in children 5 to 13. Cortex, 1973, 9, 394-402.

Bever, T. G. The nature of cerebral dominance in the speech behavior of the child and adult. In R. Huxley & E. Ingram (Eds.), Language acquisition: Models and methods. London: Academic Press, 1971.

Blumstein, S., & Cooper, W. E. Hemispheric processing of intonation contours. Cortex, 1974, 10, 146-158.

Bogen, F. E. The other side of the brain II: An appositional mind. Bulletin of the Los Angeles Neurological Society, 1969, 34, 135-162.

Branch, C., Milner, B., & Rasmussen, T. Intracarotid sodium amytal for the lateralization of cerebral speech dominance. Journal of Neurosurgery, 1964, 21, 399-405.

Bryden, M. P., & Allard, F. Dichotic listening and the development of linguistic processes. In M. Kinsbourne (Ed.), Hemispheric asymmetries of function. Cambridge, England: Cambridge University Press, 1977.

Byers, R. K., & McLean, W. T. Etiology and course of certain hemiplegias with aphasia in childhood. Pediatrics, 1962, 29, 376-383.

Carmon, A., & Bechtoldt, H. P. Dominance of the right cerebral hemisphere for stereopsis. Neuropsychologia, 1969, 7, 29-39.

Carmon, A., & Benton, A. L. Tactile perception of direction and number in patients with unilateral cerebral disease. Neurology, 1969, 19, 525-532.

Cohen, G. Hemispheric differences in serial versus parallel processing. Journal of Experimental Psychology, 1973, 3, 349-356.

Condon, W. S., & Sanders, L. W. Synchrony demonstrated between movements of the neonate and adult speech. Child Development, 1974, 45, 456-462.

Cutting, J. E. Two left hemisphere mechanisms in speech perception. Perception & Psychophysics, 1974, 16, 601-612.

Cutting, J. E., & Eimas, P. D. Phonetic feature analyzers and the processing of speech in infants. Haskins Laboratories Status Report on Speech Research, 1974, SR-37/38, 45-63.

Cutting, J. E., & Eimas, P. D. Phonetic feature analyzers and the processing of speech in infants. In J. F. Kavanaugh & J. E. Cutting (Eds.), The role of speech in language. Cambridge, Mass.: M.I.T. Press, 1975.

Dennis, M., & Kohn, B. Comprehension of syntax in infantile hemiplegics after cerebral hemidecortication: Left hemisphere superiority. Brain & Language, 1975, 2, 472-482.

Dennis, M., & Whitaker, H. A. Language acquisition following hemidecortication: Linguistic superiority of the left over the right hemisphere.

Brain & Language, 1976, *3*, 404-433.

Deutsch, D. Musical illusions. *Scientific American*, 1975, *233*, 92-103.

Dimond, S. J., & Beaumont, J. B. (Eds.). *Hemispheric function in the human brain*. New York: Wiley-Halsted Press, 1974.

Dorman, M. F., Studdert-Kennedy, M., & Raphael, L. J. Stop consonant recognition: Release bursts and formant transitions as functionally-equivalent, context-dependent cues. *Perception & Psychophysics*, 1977, *22*(2), 109-122.

Eimas, P. D. Auditory and linguistic processing of cues for place of articulation by infants. *Perception & Psychophysics*, 1974, *16*, 513-521. (a)

Eimas, P. D. Linguistic processing of speech by infants. In R. L. Schiefelbusch & L. L. Lloyd (Eds.), *Language perspectives -- Acquisition, retardation and intervention*. Baltimore: University Park Press, 1974. (b)

Eimas, P. D. Speech perception in early infancy. In L. B. Cohen & P. Salapatek (Eds.), *Infant perception: From sensation to cognition.* (Volume II. Perception of space, speech and sound). New York: Academic Press, 1975.

Eimas, P. D., Siqueland, E., Jusczyk, P., & Vigorito, J. Speech perception in infants. *Science*, 1971, *171*, 303-306.

Eisenberg, R. B. Auditory behavior in the human neonate: Methodological problems and the logical design of research procedures. *Journal of Auditory Research*, 1965, *5*, 159-177.

Eisenberg, R. B. Auditory behavior in the human neonate: Functional properties of sound and their ontogenetic implications. *International Audiology*, 1968, *8*, 34-45.

Emde, R. N., & Robinson, J. The first two months: Recent research in developmental psychobiology and the changing view of the newborn. In J. Noshpitz & J. Call (Eds.), *Basic handbook of child psychiatry*. New York: Basic Books, 1976.

Entus, A. K. Hemispheric asymmetry in processing of dichotically presented speech and nonspeech sounds by infants. In S. Segalowitz & F. A. Gruber (Eds.), *Language development and neurological theory*. New York: Academic Press, 1977.

Fitch, H. L., Halwes, T., Erickson, D. M., & Liberman, A. M. Perceptual equivalence of two acoustic cues for stop-consonant manner. *Haskins Laboratories Status Report on Speech Research*, 1979, *SR-57*, 161-182.

Fontenot, D. J., & Benton, A. L. Perception of direction of line in the right and left visual fields. *Neuropsychologia*, 1972, *10*, 447-452.

Geschwind, N., & Levitsky, W. Human brain: Left-right asymmetries in temporal speech region. *Science*, 1968, *161*, 186-187.

Glanville, B. B., Best, C. T., & Levenson, R. A cardiac measure of cerebral asymmetries in infant auditory perception. *Developmental Psychology*, 1977, *13*, 54-59.

Goodglass, H., & Quadfasel, F. A. Language laterality in left-handed aphasics. *Brain*, 1954, *77*, 521-548.

Gordon, H. W. Hemispheric asymmetries in the recognition of musical chords. *Cortex*, 1970, *6*, 387-398.

Graham, F. K., Berg, K. M., Berg, W. K., Jackson, J. C., & Kantowitz, S. R. Cardiac orienting responses as a function of age. *Psychonomic Science*, 1970, *19*, 363-364.

Graham, F. K., & Jackson, J. C. Arousal systems and infant heart rate responses. In L. P. Lipsitt & H. W. Reese (Eds.), *Advances in child development and behavior* (Vol. 5). New York: Academic Press, 1970.

Grey, J. M., & Gordon, J. W. Perceptual aspects of spectral modifications on musical timbres. *Journal of the Acoustical Society of America*, 1978, *63*(5), 1493-1500.

Hale, G. A. On the use of ANOVA in developmental research. *Child Development*, 1977, *48*, 1101-1106.

Halperin, Y., Nachshon, I., & Carmon, A. Shift in ear superiority in dichotic listening to temporal pattern nonverbal stimuli. *Journal of the Acoustical Society of America*, 1973, *53*, 46-50.

Hécaen, H. Acquired aphasia in children and the ontogenesis of hemispheric functional specialization. *Brain & Language*, 1976, *3*, 114-134.

Hebb, D. O. Intelligence in man after large removals of cerebral tissue: Defects following right temporal lobectomy. *Journal of General Psychology*, 1939, *21*, 437-446.

Hutt, S. J., Hutt, C., Lenard, H. G., Bernuth, H., & Muntjewerff, W. J. Auditory responsivity in the human neonate. *Nature*, 1968, *218*, 888-890.

Inglis, J., & Sykes, D. H. Some sources of variation in dichotic listening performance in children. *Journal of Experimental Psychology*, 1967, *5*, 480-488.

Kimura, D. Cerebral dominance and the perception of verbal stimuli. *Canadian Journal of Psychology*, 1961, *15*, 166-171.

Kimura, D. Speech lateralization in young children as determined by an auditory test. *Journal of Comparative & Physiological Psychology*, 1963, *56*, 899-901.

Kimura, D. Left-right differences in the perception of melodies. *Quarterly Journal of Experimental Psychology*, 1964, *14*, 355-358.

Kimura, D. Functional asymmetry of the brain in dichotic listening. *Cortex*, 1967, *3*, 163-178.

Kimura, D. Spatial localization in left and right visual fields. *Journal of Psychology*, 1969, *23*, 445-458.

Kimura, D., & Folb, S. Neural processing of backward speech sounds. *Science*, 1968, *161*, 395-396.

Kinsbourne, M. The control of attention by interaction between the hemispheres. In S. Kornblum (Ed.), *Attention and performance* (IV). New York: Academic Press, 1975.

Kirk, R. E. *Experimental design: Procedures for the behavioral sciences*. Belmont, Calif.: Wadsworth Publishing Co., 1968.

Knox, A. W., & Kimura, D. Cerebral processing of nonverbal sounds in boys and girls. *Neuropsychologia*, 1970, *8*, 227-238.

Kohn, B., & Dennis, M. Selective impairments of visuo-spatial abilities in infantile hemiplegics after right cerebral decortication. *Neuropsychologia*, 1974, *12*, 505-512.

Lenneberg, E. H. *Biological foundations of language*. New York: John Wiley & Sons, 1967.

Lenneberg, E. H. On explaining language. *Science*, 1969, *164*, 635-643.

Levy, J. Psychobiological implications of bilateral asymmetry. In S. J. Dimond & J. G. Beaumont (Eds.), *Hemispheric function in the human brain*. New York: Wiley-Halsted Press, 1974.

Levy-Agresti, J., & Sperry, R. Differential perceptual capacities in major and minor hemispheres. *Proceedings of the National Academy of Sciences*, 1968, 61, 1151.

Liberman, A. M. The grammars of speech and language. Cognitive Psychology, 1970, 1, 301-323.

McAdam, D. W., & Whitaker, H. A. Language production: Electroencephalographic localization in the normal human brain. Science, 1971, 172, 499-502.

McCall, R. B., & Melson, W. H. Amount of short term familiarization and the response to auditory discrepancies. Child Development, 1970, 41, 861-869.

McFie, J. Intellectual impairment in children with localised post-infantile cerebral lesions. Journal of Neurology, Neurosurgery & Psychiatry, 1961, 24, 361-35.

McFie, J., & Thompson, J. A. Variation with age of the effects of cerebral lesions in man. Brain Research, 1971, 31, 363.

McKee, G., Humphrey, B., & McAdam, D. W. Scaled lateralization of alpha activity during linguistic and musical tasks. Psychophysiology, 1973, 10, 441-443.

Miller, C. L., & Morse, P. A. The "heart" of categorical speech discrimination in young infants. Journal of Speech and Hearing Research, 1976, 19, 578-589.

Miller, C. L., Morse, P. A., & Dorman, M. F. Cardiac indices of infant speech perception: Orienting and burst discrimination. Quarterly Journal of Experimental Psychology, 1977, 29, 533-545.

Milner, B. Laterality effects in audition. In V. B. Mountcastle (Ed.), Hemispheric relations and cerebral dominance. Baltimore: Johns Hopkins Press, 1962.

Milner, B. Hemispheric specialization. In F. O. Schmidt & F. G. Worden (Eds.), The neurosciences: Third study program. Cambridge, Mass.: M.I.T. Press, 1974.

Mirabile, P. J., Porter, R. J., Hughes, L. F., & Berlin, C. I. Dichotic lag effect in children 7 to 15. Developmental Psychology, 1978, 14, 277-285.

Moffitt, A. R. Consonant cue perception by twenty- to twenty-four week old infants. Child Development, 1971, 42, 717-731.

Molfese, D. L., Freeman, R. B., & Palermo, D. S. The ontogeny of brain lateralization for speech and nonspeech stimuli. Brain & Language, 1975, 2, 356-368.

Molfese, D. L., Nunez, V., Seibert, S. M., & Ramaniah, N. V. Cerebral asymmetry: Changes in factors affecting its development. Annals of the New York Academy of Sciences, 1976, 280, 821-833.

Morse, P. A. The discrimination of speech and nonspeech stimuli in early infancy. Journal of Experimental Child Psychology, 1972, 14, 477-492.

Morse, P. A. Infant speech perception: A preliminary model and review of the literature. In R. L. Schiefelbusch & L. L. Lloyd (Eds.), Language perspectives -- Acquisition, retardation and intervention. Baltimore: Universty Park Press, 1974.

Morse, P. A. Infant speech perception: Origins, processes, and Alpha Centauri. Infant Development Laboratory: Research Status Report, 1978, 2, 89-144.

Natale, M. Perception of nonlinguistic auditory rhythms by the speech hemisphere. Brain & Language, 1977, 4, 32-44.

Nebes, R. D. Superiority of the minor hemisphere in commissurotomized man for the perception of part-whole relations. Cortex, 1970, 7, 333-349.

Petersen, M. R., Beecher, M. D., Zoloth, S. R., Moody, D. B., & Stebbins,

W. C. Neural lateralization of species-specific vocalizations by Japanese macaques (*Macaca fuscata*). *Science*, 1978, *202*, 324.

Porter, R. J. Jr., & Berlin, C. I. On interpreting developmental changes in the dichotic right-ear advantage. *Brain & Language*, 1975, *2*, 186-200.

Rasmussen, T., & Milner, B. The rôle of early left-brain injury in determining lateralization of cerebral speech functions. *Annals of the New York Academy of Sciences*, 1977, *299*, 355-369.

Satz, P., Bakker, D. J., Teunissen, J., Goebel, R., & Van der Vlugt, H. Developmental parameters of ear asymmetry: A multivariate approach. *Brain & Language*, 1975, *2*, 171-185.

Shankweiler, D. Effects of temporal lobe damage on perception of dichotically presented melodies. *Journal of Comparative & Physiological Psychology*, 1966, *62*, 115-119.

Shankweiler, D., & Studdert-Kennedy, M. Identification of consonants and vowels presented to the left and right ears. *Quarterly Journal of Experimental Psychology*, 1967, *19*, 59-63.

Simon, N. Echolalic speech in childhood autism: Consideration of possible underlying loci of brain damage. *Archives of General Psychiatry*, 1975, *32*(11), 1439-1446.

Studdert-Kennedy, M. Speech perception. In N. J. Lass (Ed.), *Contemporary issues in experimental phonetics*. New York: Academic Press, 1976.

Studdert-Kennedy, M., & Shankweiler, D. Hemispheric specialization for speech perception. *Journal of the Acoustical Society of America*, 1970, *48*, 579-594.

Takagi, R. [Language developmental disorder as the primary disturbance of childhood autism.] *Japanese Journal of Child Psychiatry*, 1972, *13(5)*, 285-294.

Tanguay, P. E. Clinical and electrophysiological research. In E. R. Ritvo, B. J. Freeman, E. M. Ornitz, & P. E. Tanguay (Eds.), *Autism: Diagnosis, current research and management*. New York: Spectrum Publications, 1976.

Teuber, H. L. Mental retardation after early trauma to the brain. In C. R. Angle & E. A. Bering, Jr. (Eds.), *Physical trauma as an etiologial element in mental retardation*. Washington, D.C.: National Institutes of Health Publications, 1970.

Trehub, S. E., & Rabinovitch, M. S. Auditory-linguistic sensitivity in early childhood. *Developmental Psychology*, 1972, *6*, 74-77.

Wada, J. A., Clark, R., & Hamm, A. Cerebral asymmetry in humans: Cortical speech zones in 100 adult and 100 infant brains. *Archives of Neurology*, 1975, *32*, 239-246.

Webster, W. G., & Webster, I. H. Anatomical asymmetry of the cerebral hemispheres of the cat brain. *Physiology & Behavior*, 1975, *14*, 867-869.

Witelson, S., & Pallie, W. Left hemisphere specialization for language in the newborn: Neuroanatomical evidence for asymmetry. *Brain*, 1973, *96*, 641-646.

Zatorre, R. J. Recognition of dichotic melodies by musicians and nonmusicians. *Neuropsychologia*, 1979, in press.

Zoloth, S. R., Petersen, M. R., Beecher, M. D., Green, S., Marler, P., Moody, D. B., & Stebbins, W. Species-specific perceptual processing of vocal sounds by monkeys. *Science*, 1979, *204*, 870-873.

FOOTNOTES

1. Although the claim that neocortical functional lateralization is uniquely human (or perhaps shared only by humans and songbirds) is still widely accepted, a small but important amount of recent evidence calls that claim into question. Research on cat visual perception and visual neuroanatomy in one laboratory suggests a right hemisphere advantage for visual pattern discrimination in that species (e.g., Webster & Webster, 1975), while work at another laboratory suggests there may be a feline left hemisphere predominance in auditory evoked responses to human speech (Molfese et al., 1976). Of even greater significance is the very recent finding of a monaural right-ear advantage in Japanese macaques for making categorizations of their species-specific vocalizations based on communicatively-relevant fine-grain spectral details of the calls (Petersen, Beecher, Zoloth, Moody, & Stebbins, 1978). Furthermore, only that species of macaque shows the REA for those calls, and only they show categorical perception (akin to human categorical phonetic perception) for communicatively-relevant aspects of the calls (Zoloth, Petersen, Beecher, Green, Marler, Moody, & Stebbins, 1979). That functional brain lateralization may *not* be uniquely human puts a different light on questions of its evolution in humans, but does not necessarily dilute its importance in human behavior (in fact, it probably accentuates it by placing it appropriately within a broader ethological framework of communicative behavior).

2. Testing ran from August, 1976, to June, 1977, during which time outside temperatures ranged from 30° below zero to +100° F. Laboratory temperatures, however, ran consistently around +70° F, and the infants were acclimated to the laboratory for about 20 minutes before testing began.

3. Loss rate within subgroups ranged from 38.46% for 2-month-old boys to 71.43% for 3-month-old girls and 4-month-old boys. Of the 48 infants who completed the study, 45.8% were first-borns. The participants' parents were young (mother's age: *M* = 25.3 years, *S.D.* = 3.7; father's age: *M* = 27.2 years, *S.D.* = 4.2), and fairly well-educated (mother's education: *M* = 17.2 years, 66.7% had some college; father's education: *M* = 14.2 years, 58.3% had some college).

4. Both the speech and music stimuli were the same as those used in the Glanville et al. (1977; Best & Glanville, Note 9) study.

5. Although the time course of habituation to dichotic speech was not reliably different from that for dichotic music notes, the infants did show a *general* difference in response magnitude to speech versus music. This finding corroborates earlier evidence that speech is especially effective in gaining the attention of, pacifying, and otherwise changing the ongoing activity patterns of young infants (e.g., Condon & Sanders, 1974; Eisenberg, 1965, 1969; Hutt, Hutt, Lenard, Bernuth, & Muntjewerff, 1968). A significant Stimulus Type effect, $F_{(1,42)} = 7.83$, $p < .007$, in the magnitude of the heart rate difference scores throughout the habituation trials indicated that cardiac ORs were larger for speech presentations than for music note presentations. This finding is further supported by a Stimulus Type X Period interaction, $F_{(1,42)} = 6.63$, $p < .025$, in the Period analysis. Simple effects tests found that whereas prestimulus heart rates did not differ significantly

for the two stimulus types, poststimulus heart rate was lower for speech syllables than for music notes, $F_{(1,84)} = 4.6476$, $p < .05$, upholding the interpretation that speech produced a larger OR than music. In spite of this stimulus difference in the overall magnitude of the OR, the prestimulus versus poststimulus heart rate difference was significant (reliable cardiac OR) for both music note presentations, $F_{(1,84)} = 13.918$, $p < .001$, and speech presentations, $F_{(1,84)} = 38.439$, $p < .0001$, during the habituation trials. As neither the Stimulus Type X Trial nor the Stimulus Type X Trial X Period interactions were significant, there is no evidence of stimulus differences in the rate or form of OR habituation.

6. Age differences in cardiac response to the dichotic stimuli during the habituation trials are reflected in a significant Age effect in the difference score analysis, $F_{(1,42)} = 14.34$, $p < .005$, and an Age X Period interaction in the Period analysis, $F_{(2,42)} = 3.81$, $p < .05$. The data indicate that tonic heart rate level decreased with age, whereas the magnitude of the stimulus-evoked cardiac OR increased with age. These findings are consonant with those of studies on the ontogeny of the cardiac response to sensory stimulation in early infancy (e.g., Graham, Berg, Berg, Jackson, & Kantowitz, 1970; Graham & Jackson, 1970). Simple effects tests comparing prestimulus heart rate with poststimulus rate for each age group showed that, on the average, across habituation trials the Period difference was not significant for the 2-month-olds, while poststimulus rate was lower than prestimulus rate for the 3-month-olds, $F_{(1,42)} = 10.15$, $p < .05$, and the 4-month-olds, $F_{(1,42)} = 11.902$, $p < .025$. Both prestimulus and poststimulus heart rate levels differed significantly among the three age groups, $F_{(2,84)} = 3.084$, and $F_{(2,84)} = 3.97$, $p < .025$, respectively. Planned comparison paired t-tests on the average heart rates for each group revealed a significant difference in average rate between 2- and 3-month-olds, $t = 10.59$, $p < .05$, and between 2- and 4-month-olds, $t = 17.85$, $p < .01$, but not between 3- and 4-month-olds. Age differences in infant tonic heart rate and in magnitude of the cardiac OR may begin to asymptote around 3 to 4 months of age, at least in response to intermittent dichotic auditory stimulation.

7. Eight 2-month-olds provided evidence for discrimination of the music change on trial 10, 6 of whom displayed an LEA. Nine infants in that age group showed speech discrimination, 6 of whom showed an REA. The magnitude of the dishabituation OR on the speech tests in these 2-month-olds was small, however, as would be expected based on the reported parametric analyses. In the 3-month group, 15 infants showed some music timbre discrimination, with an LEA in 10 of these; 13 infants showed some speech discrimination, with an REA in 12 infants and an equal discrimination by both ears in one infant. Of the 4-month-olds, 10 showed some music timbre discrimination, 6 of whom had an LEA; 11 showed some speech discrimination, of whom 6 had an REA. However, the magnitude of the OR recovery to the speech change was greater for those 4-month-olds showing an REA than for those showing an LEA, accounting for the results of the parametric analyses.

CROSS-SERIES ADAPTATION USING SONG AND STRING

Robert E. Remez, James E. Cutting,[++] and Michael Studdert-Kennedy[+++]

Abstract. The acoustic-auditory feature "rise-time" has been claimed to underlie both the phonetic *affricate*-*fricative* distinction and the nonphonetic *plucked* *string*-*bowed* *string* distinction. We used the perceptual adaptation technique to determine whether the rise-time differences of the [dʒa]-[ʒa] distinction would therefore be registered by the same mechanism that mediates rise-time differences for the plucked-bowed distinction. Two continua were used, one of digitally-modified natural speech and one of synthetic violin sounds, in which the rise-time was varied across each set of tokens from 0 msec to 80 msec in steps of 10 msec. The speech was sung and the violin notes were synthesized with the same fundamental frequency, 294 Hz. Adaptation of the category boundaries was observed only when speech adaptors were tested with the speech continuum and when violin adaptors were tested with the violin continuum. When cross-series tests were performed (violin adaptors tested with the speech series, and speech adaptors tested with the violin series), no effect of adaptation was observed. This finding indicates that these speech and violin sounds, despite obvious acoustic similarities, do not share the same feature detectors.

INTRODUCTION

Is speech perception merely an auditory process? The discovery that different modes of sound production of the violin are perceived categorically (Cutting & Rosner, 1974) seemed to suggest that it might be. Categorical perception had previously been thought to result from a decoding process of the speech perception system (Liberman, Cooper, Shankweiler, & Studdert-Kennedy, 1967). Phonetic decoding of acoustic signals was believed to be abstractly keyed to articulatory patterns of phonetic production (Cooper, Liberman, Harris, & Grubb, 1958). In the earliest form of the hypothesis, speech segments produced by graded motor signals (e.g., vowels) were said to be perceived in a continuous fashion, in contrast to those produced by discrete motor commands (e.g., consonants), which were said to be perceived categorically. Though subsequent study of articulator coordination and motor

---

[+] Also Indiana University.

[++] Also Wesleyan University.

[+++] Also Queens College and Graduate Center, City University of New York.

Acknowledgment: This work was supported by NICHD Grant HD-01994, and by a grant to the first author from NIMH, Grant MH 32848-01. We are grateful to Linda B. Smith for her advice and comments on the manuscript.

[HASKINS LABORATORIES: Status Report on Speech Research SR-59/60 (1979)]

control ruled out literal parity of production and perception (MacNeilage, 1970), categorical perception continued to be an important criterion for separating speech perception from general auditory perception (Liberman, 1970). However, when Cutting and Rosner (1974) reported that a stimulus continuum ranging from plucked string to bowed string was categorically perceived, they clearly eliminated categorical perception as one of the definitive qualities of perception in the speech mode. This result, in combination with other similar findings (e.g., Locke & Kellar, 1973; Miller, Weir, Pastore, Kelly, & Dooling, 1976; Pisoni, 1977), intensified the appeal of explaining speech perception by means of general auditory (lower level) mechanisms.

A conservative account of the categorical perception of violin-made sounds was that the distinction between plucked string and bowed string was actually mediated by mechanisms intrinsic to the speech system. Cutting and Rosner (1974) considered this possibility because they saw that the acoustic basis for perceiving different violin articulations is very nearly identical to that of the phonetic affricate/fricative distiction (Gerstman, 1957). Both plucked/bowed and affricate/fricative may be distinguished by the rate of onset of acoustic energy, the rise-time of the amplitude envelope. When rise-time is relatively short--less than 40 msec--listeners hear a plucked string or an affricate consonant; when rise-time is long--greater than 40 msec--listeners hear a bowed string or a fricative. However, Cutting and Rosner did not defend this proposal. Instead, they favored a general auditory mechanism that might serve both phonetic and violin distinctions by tracking amplitude changes in the acoustic signal. This auditory explanation seemed more plausible than the proposal that the speech system had mistakenly processed the synthetic violin sounds categorically, primarily because it was not obvious why the perceptual mechanism would misapply the phonetic code to these particular nonphonetic patterns. In addition, such an explanation in the case of violins would be inconsistent with supposed biological advantages of categorical perception: If the biologically-determined speech code is unique among the acoustic patterns that engage the auditory system, then the speech processor has a very easy job of telling speech sounds from nonspeech (Liberman, Mattingly, & Turvey, 1972); it must merely listen for stimuli of requisite encodedness. From this standpoint it would be inappropriate to argue that the stable, efficient speech system would be easily fooled by violin notes. In short, Cutting and Rosner concluded that the perception of violin articulation is probably not accomplished by the speech perception mechanism.

The goal of Cutting, Rosner, and Foard (1976), then, was to determine a possible sensory basis of categorical perception (as opposed to the motoric rationale offered by Liberman et al., 1967); they took the pattern recognition scheme of opponent-related feature detectors introduced by Eimas and Corbit (1973) as their model. But, whereas Eimas had originally conceived of the detectors as phonetic, Cutting et al. saw that auditory detectors would be required to serve parsimoniously for both speech and violin sounds. Had the violin sounds not been categorically perceived, the phonetic detector model could have trivially accommodated nonspeech phenomena at a lower, auditory level. However, since the plucked/bowed distinction was categorical, it appeared that an explanation of the general phenomenon of categorical perception should be recast nonphonetically, in more general auditory terms. To the degree that adaptation experiments establish detector sensitivities (cf. Remez, 1979), the demonstration by Cutting et al. of adaptation effects on identification functions for synthetic violin sounds argued that categorical perception of both speech and nonspeech sounds might be mediated by the same set of auditory detectors.

Cutting et al. used their violin test series in two adaptation conditions; both conditions revealed significant adaptation effects. In the *within-series* condition the adaptors were drawn from the test continuum; the adaptor was either the plucked endpoint or the bowed endpoint. In *selective* adaptation tests, adaptors were not drawn from the continuum but were fashioned to fatigue the detectors for specific attributes of the test stimuli. Identification performance in these cases was affected by adaptors from the test series differing in either waveform (sinusoid, sawtooth) or frequency independently, although boundary shifts were diminished relative to within-series shifts. Although the experiment did not directly test the hypothesis that a single auditory mechanism sensitive to rise-time was responsible for the analysis of plucked/bowed strings and affricate/fricative consonants, it did reveal that a style of perceptual analysis similar to that proposed for speech appeared to operate at an auditory level for nonspeech sounds. The implication to be drawn from the adaptation results of Cutting et al. (1976) is clear: Auditory analyzers alone might suffice for the categorical perception of both speech and violin sounds.

The present experiment uses the adaptation paradigm in a direct test of the hypothesis that a single set of auditory analyzers tuned to amplitude rise-time serves for both speech and nonspeech perception. To perform this test, two stimulus series were fashioned, one of synthetic plucked-to-bowed violin sounds, the other of computer-modified natural affricate-to-fricative consonant-vowel syllables. These series were as similar as possible in the onset portions critical to the respective perceptual distinctions. In the speech case, the continuum ranged from [dʒa] (affricate) to [ʒa] (fricative), after Gerstman (1957); in the violin series, the synthetic continuum ranged from the sound of a plucked string to that of a bowed string (after Cutting & Rosner, 1974). Our test consisted of two parts: In the first, adaptation caused by consonant endpoint or violin endpoint adaptors was measured by using the stimulus test series from which the adaptor was drawn. This was the *within-series* test. In the second, adaptation caused by speech adaptors was measured with the violin test series, and adaptation caused by violin adaptors was measured with the speech series. This was the *cross-series* test. Our purpose was to establish whether an amplitude rise-time detector, fatigued by a nonspeech stimulus, would effect a perceptual change in a speech test series, and vice versa. The demonstration of cross-series adaptation in this circumstance would be strong evidence for a common analytic mechanism underlying the categorical perception of speech and nonspeech sounds.

METHOD

Subjects

Eight volunteers served as listeners. All were undergraduate students enrolled at the University of Connecticut. They were right-handed native speakers of English; none had a history of speech disorder or hearing

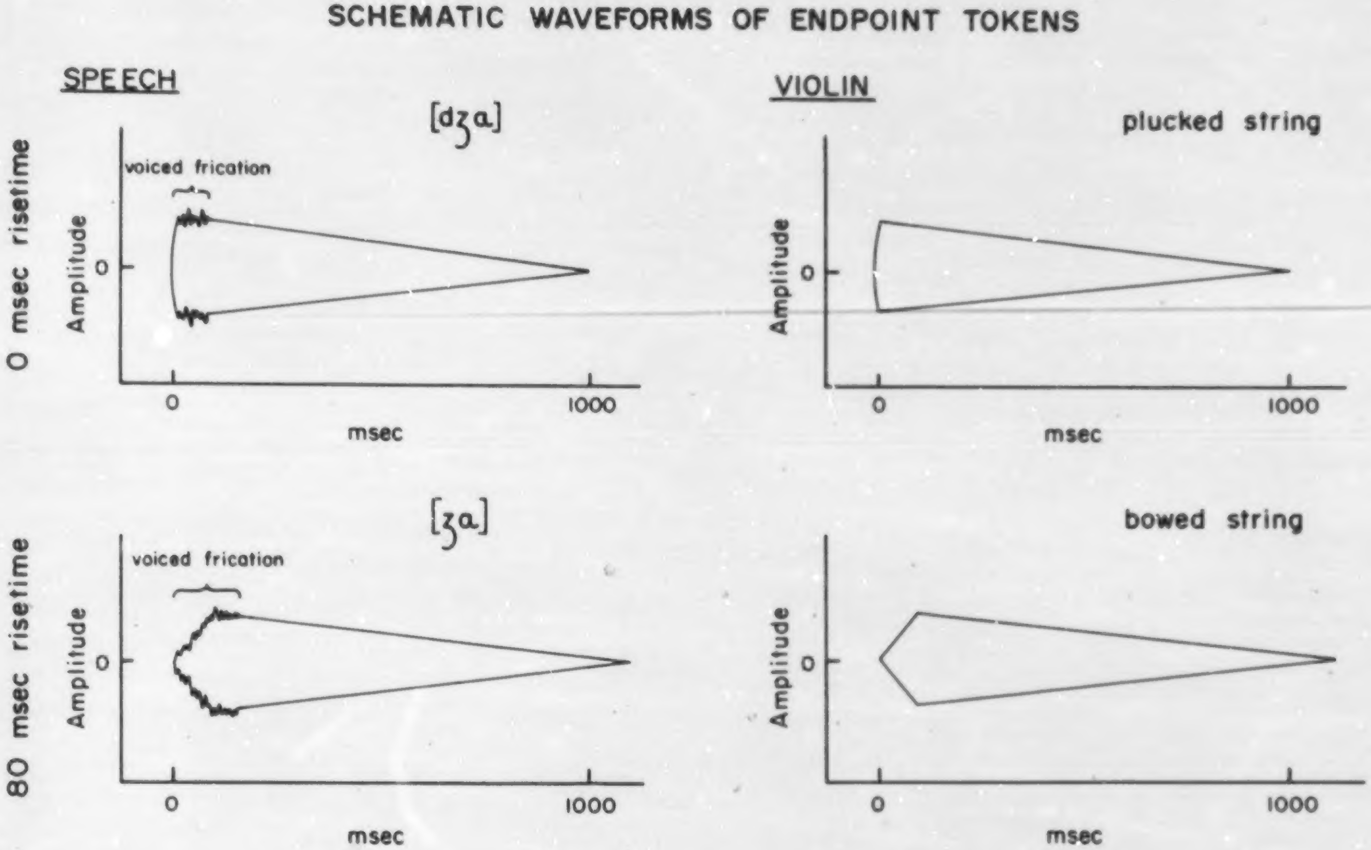


Figure 1. Schematic waveforms of endpoint tokens. Left: speech endpoints, [dʒa] (top), and [ʒa] (bottom). Right: violin endpoints, plucked string (top), bowed string (bottom)

impairment, nor did any confess more than a casual familiarity with musical instruments of the violin family. They were paid for their participation.

Stimuli

Two stimulus continua were employed, one of synthetic violin sounds and the other of computer-modified natural speech. The violin series was one of the nine-item sets of Cutting and Rosner (1974), in which the amplitude rise-time of a 294 Hz sawtooth wave was varied from 0 msec to 80 msec in steps of 10 msec. Overall duration varied from 1020 msec to 1100 msec. The matching speech series was generated from a CV syllable, [ʒa], sung by one of the authors (J.E.C.) in falsetto register at 294 Hz. The syllable was tape-recorded, low-pass filtered at 5 kHz, sampled at 10 kHz, and digitized on the Haskins Laboratories pulse-code modulation system (Cooper & Mattingly, 1968). The amplitude rise-time was then shaped by iterative multiplication of the digital record in small steps to produce a series of sounds that ranged from [dʒa] to [ʒa]. The nine-item continuum varied from 0 msec to 80 msec of rise-time; the duration of the voiced frication in the initial portion of each token covaried with rise-time from 70 msec to 150 msec (after Gerstman, 1957). Schematic waveforms of the four endpoints are presented in Figure 1. Test sequences were recorded on tape by digital-to-analog conversion and were presented binaurally to listeners via Crown 820-144 playback through TDH-39 earphones at a comfortable level of about 72 dB SPL.

Procedure

Eight tests were presented on separate days over the course of two weeks; four were of within-series adaptation, four of cross-series. The endpoint tokens of the two continua were used as adaptors. In the within-series conditions, the test continuum and adaptor were of the same kind, both either speech or violin. In cross-series tests, the adaptor was of one kind, and the test series of the other. Subjects were tested in two groups of four; one group was tested on within-series first, cross-series second; the other group was tested in the reverse order.

Brief practice with four sequential presentations of each day's test series endpoints began the testing session. Following this was a baseline test of unadapted identification of the relevant stimulus series, a randomized 90-item sequence of the nine tokens repeated ten times each. Three seconds of silence separated trials. After a short intermission, adapted identification was measured, following the procedure of Cutting et al. (1976). This test consisted of twelve blocks of trials of alternating adaptor and identification portions. In the first block there were 100 adaptor repetitions followed by seven identification trials drawn from the nine-item continuum. The remaining eleven blocks were similar to the first but with only 50 adaptor repetitions. There were 600 msec of silence between successive adaptors; 2 sec between adaptor block and identification block; 2 sec between successive identification trials; and 5 sec between the end of the identification block and the following adaptor block. Six observations per listener were obtained for items with rise-times of 0, 10, 70, and 80 msec, and twelve observations for those with rise-times of 20, 30, 40, 50, and 60 msec. Responses were written on a prepared answer sheet as "P" (plucked), "B" (bowed), "J" ([dʒa]), or "Z" ([ʒa]).

RESULTS AND DISCUSSION

For each within- and cross-series adaptation test we calculated two scores per subject: one for baseline identification, one for adapted identification. In each of the eight conditions, the proportion of "long rise" responses ([ʒa] or bowed) was compared in baseline and adapted tests by a one-tailed t-test for paired observations. Within-series differences between baseline and adapted identification were statistically significant but there were no significant differences between baseline and adapted identification in cross-series tests (violin adaptors tested with the speech series, and speech adaptors tested with the violin series). Table 1 summarizes the results.

Table 1

Mean within-series and cross-series adaptation effects, scored as differences in the proportions of long-rise responses.

Condition	Baseline minus Adaptation	t	
[dʒa] on speech	.122	4.798	$p<.005$
[ʒa] on speech	-.089	3.927	$p<.005$
plucked on violin	.052	3.260	$p<.01$
bowed on violin	-.097	4.140	$p<.005$
[dʒa] on violin	.002	0.072	$p>.1$
[ʒa] on violin	-.016	0.588	$p>.1$
plucked on speech	.002	0.154	$p>.1$
bowed on speech	.007	0.364	$p>.1$

Our experiment reveals that test continuum and adaptor must share more than the mere acoustic attributes of amplitude envelope and fundamental frequency to show cross-series adaptation effects. This is consistent in principle with the prior finding of Cutting et al. (1976) that in cross-series adaptation the magnitude of the effect was a function of the auditory attributes that the test series shared with the adaptor. However, since the attribute of rise-time was always common to adaptor and test series in their study, we had supposed that sounds sharing amplitude rise-time might be mutually effective adaptors, regardless of the other auditory attributes distinguishing them. Rise-time, as a perceptual commodity, might then have been parsimoniously assigned to a single pair of detectors.

The outcome of the present experiment refutes this hypothesis. In our test, cross-series adaptors and test continua differed neither in amplitude rise-time nor in fundamental frequency, but did differ in spectral composition, and this spectral difference (or in other words, the apparent difference in sound-source) was sufficient to prevent cross-adaptation of the two series. Evidently, speech and violin sounds do not share a single 0 msec-80 msec pair of rise-time detectors. By extrapolation, we may conclude that the categorical perception of the affricate/fricative and plucked/bowed distinctions does not rely on the detection of rise-time by a single auditory pattern analyzer.

In this light, an earlier report by Diehl (1976) is puzzling. Diehl tested the generality of the *stop* versus *continuant* opposition using nonspeech sounds as adaptors. He found that a plucked adaptor was effective on a [ba]-[wa] continuum, as if it shared the property "stop" with [ba], although a similar test using a bowed adaptor showed no adaptation effect. What is even more curious is that the [ba] and the plucked sound, which had similar effects as adaptors, were not similar in amplitude envelope, nor was the [ba]-[wa] distinction cued by rise-time. Finally, the frequency of the violin adaptors was 440 Hz, while the speech sounds presumably had a fundamental at least two octaves lower. In short, the cross-series adaptation reported by Diehl for violin and speech sounds sharing neither rise-time, nor fundamental frequency, nor spectral composition must surely be the most abstract instance of adaptation ever observed. Whatever the critical attribute may have been, it was plainly not the abtract *stop* versus *continuant* opposition, as Diehl claimed, since our [dʒa], which did not cross-adapt, is a proper stop, and [ʒa] is a proper continuant.

A related set of findings was reported recently by Samuel and Newport (1979), who replicated Diehl's earlier (1976) puzzling result. They used two test series, [ba] to [wa], and [tʃa] to [ʃa]; and four nonspeech adaptors, two periodic and two aperiodic, of short and long rise-time each. The periodic nonspeech adaptor pair were plucked and bowed violin sounds that differed in fundamental period from the speech series, again, presumably by two octaves. The aperiodic adaptor pair were broad-band noise patterns with digitally shaped envelopes; no apparent linguistic qualities were reported for the noise adaptors, e.g., they did not sound like whispered speech. Each of the four adaptors was tested on both speech continua. The surprising findings in this study were that the periodic fast rise-time adaptor acted, again, as if it shared a property with [ba] *but not with* [tʃa]; and the slow rise-time aperiodic adaptor acted as if it shared a property with [ʃa] *but not with* [wa]. Confronted by this set of asymmetrical rise-time adaptation effects, the authors posited an alteration to the original detector model of Eimas: They suggested that periodic waveforms were mediated by a single detector sensitive only to sharp envelope discontinuities, while aperiodic waveforms were mediated by a single detector sensitive only to gradual energy onset. However, one test that Samuel and Newport did not perform is crucial to supporting this conclusion. To verify the characterizations of the two *auditory* detectors, speech adaptors must also be effective when the test series is nonspeech. Only if the effects of speech adaptors on nonspeech test continua are shown to correspond to the effects of nonspeech adaptors on speech test continua would the proposal of auditory, periodicity-labelled asymmetrically-tuned rise-time detectors be warranted.

In addition, our present test failed to confirm their hypothesis. By the criteria developed by Samuel and Newport (1979), we might have predicted that only the plucked string and [dʒa] adaptors would be effective, though each should have produced adaptation on both test continua. This is because both the speech and the violin continua were periodic (at the *same* fundamental frequency, 294 Hz). The prediction was incorrect. Moreover, their conclusions are open to criticism independent of our disconfirming cross-series data. The detectors they have proposed account only for the success of plucked adaptors concomitant with the failure of bowed adaptors. Although this outcome occurred in their test, their conclusion is overly general; a periodic rise-time detector that asymmetrically prefers sharp amplitude discontinuities would not explain the original symmetrical within-series and selective rise-time adaptation effects of Cutting et al. (1976) nor the symmetrical within-series effects noted here. Our data may therefore be taken to question the usefulness of the modification of the detector model endorsed by Samuel and Newport.

GENERAL DISCUSSION

Our finding that adaptation effects are segregated by apparent sound-source has a precedent: Faced with an analogous result, of fundamental-frequency-contingent phonetic adaptation, Ades (1977) suggested that multiple sets of auditory detectors may exist, each assignable to an individual talker. In our case, we might extend the source-assignment notion to *sound*-sources in general, whether sources of phonetic segments or not. For instance, rise-time detectors might be basic ingredients in auditory detector ensembles; speech and violin notes, produced by different sound-sources, would then be mediated perceptually by separate but equal ensembles, one assigned to each sound-source. This vague notion of (re)duplicated auditory detectors will handle our result, but it is both *ad hoc* and inelegant, particularly in view of the current controversy over the status of feature detectors in speech perception.

The history and the interpretive difficulties of adaptation studies using phonetic materials have been reviewed by Cooper (1975) and by Eimas and Miller (1978). To the difficulties exposed by these authors we may add the following: (1) detectors-conceived-phonetically fail to predict the occasions on which adaptation does and *does not* occur; (2) detectors-conceived-auditorily have so proliferated as to eliminate the appealing simplicity of the original model: Cumbersome and implausible detector interactions are required to derive perceptual categories in the auditory version of the model; and, (3) the adaptation test itself may not be appropriate for cataloging fixed preferences and sensitivities of analytic channels (see Weisstein, 1964, p. 164; Simon & Studdert-Kennedy, 1978; Diehl, Elman, & McCusker, 1978; Remez, 1979). We therefore question the worth of explaining our current finding within the feature detector model.

Further, we find an independent dismissal of feature detector accounts in a recent study quite analogous to our own (Pisoni, Note 1). Pisoni observed no cross-series adaptation effects for speech and nonspeech stimulus sets that were equivalent on an acoustic dimension that is sometimes taken to be perceptually critical in speech (cf. Lisker & Abramson, 1964, 1971). The speech set was one of *voicing*, the nonspeech set one of two-component tones onset at differing degrees of asynchrony. As in the case of Cutting et al. (1976) and in the present experiment, the acoustic criteria for the speech and nonspeech distinctions were apparently the same. The acoustic parameter that served to distinguish *coincident* from *delayed* voicing onset also served to distinguish the "categories" of the nonspeech set; this was the temporal relation of the onsets of the individual spectral components. In adaptation tests, the two-tone continuum exhibited within-series adaptation, as did the voicing continuum, which is well known to undergo adaptation (e.g., Eimas & Corbit, 1973)--but neither series exhibited any cross-series effects. Thus, identical temporal structure did not guarantee cross-adaptation.

Rather than invent a new pair of detectors to handle the finding, and, rather than take the tack of Ades (1977) by inventing an entire new set of detectors assigned ad hoc to nonspeech tones, Pisoni preferred to argue that temporal order judgments of any class may be limited by a general constraint of the auditory system unspecific to particular detectors. Perception of the amplitude rise-time parameter that we have studied for violin and speech sounds may be subject to a related general auditory limitation, one not specific to any pair of phonetic or auditory detectors. If the auditory system had a characteristic rate of temporal integration throughout its frequency range, for instance, then the transduction of every sound impinging on the system would reflect that limitation in resolution. In some circumstances, we might expect a perceptual distinction to reflect differential effects of this auditory limitation. But, when the effect of the general auditory constraint happens to be perceptually distinctive, we should not assume that the mechanism responsible is a single pair of detectors or even a pair of ensembles of detectors. Were that the case, our closely matched set of speech and violin sounds would have adapted each other, for they were as similar as possible within the requirement that they differ in their apparent sources, one a human talker, the other a violin.

Finally, we may consider the possible origins of the violin perceptual categories of which the time courses happen to coincide with those of certain speech sounds. Recently, Remez (1978) has argued that violins are capable of categorical and continuous modes of sound production, though this distinction is a mechanical one, unlike the neurophysiological claims made in the case of speech (Cooper et al., 1958; Liberman et al., 1967; cf. Stevens, 1972). For example, the productive distinction sul ponticello-sul tasto is acoustically correlated with differences in spectral envelope; sul ponticello (played with the bow near the bridge) has a shallower rolloff, that is, more energy in the higher harmonics, than does sul tasto (played with the bow near the fingerboard), all other things equal (Schelleng, 1973). Given constant bow force, then, this dimension of production is continuously variable in infinitely small steps as the point of contact of bow and string is moved from the bridge toward the fingerboard, or vice versa. Compare this to the pizzicato-arco (plucked-bowed) distinction. These terms name two quantal alternatives of sound production. Bowing involves what Schelleng (1973) called the "stick-slip" interaction of string and bow in which force is applied to the string in a relatively sustained manner. Plucking, on the other hand, involves a single loading of the string when it is retracted by a finger or other grasping agent (such as a plectrum in a harpsichord). Once loaded, the string is released. The differences in production are correlated acoustically with the rise-time differences we have been discussing. Unlike the gradual change from sul ponticello to sul tasto, the distinction between pizzicato and arco is categorical productively; there is no mechanical gradient from plucking a string to bowing it. Listeners who perceive the mechanical events of instrumental articulation, by this line of reasoning, ought to perceive the plucked/bowed distinction categorically. Perhaps the finding of violin categoricity that has inspired the studies by Cutting and his colleagues is as much a function of the perception of intrinsically categorical articulatory events as it is a function of the auditory resolution of rise-time.

To summarize, our experiment used a cross-series adaptation test to determine whether rise-time of amplitude envelope is mediated by a single auditory mechanism in both phonetic and nonphonetic circumstances. Although we found within-series effects for speech and violin sounds, cross-series tests did not reveal adaptation. Thus, stimuli matched for fundamental frequency, amplitude envelope and rise-time apparently do not share feature detectors--if we judge by the conventional criteria for establishing shared detectors. We may further conclude that the categorical perception of violin articulation does not depend on the same auditory mechanism as is used for the phonetic distinction between affricate and fricative.

REFERENCE NOTE

1. Pisoni, D. B. *Adaptation of the relative onset time of two-component tones*. Unpublished manuscript, 1979.

REFERENCES

Ades, A. E. Source assignment and feature extraction in speech. *Journal of Experimental Psychology: Human Perception and Performance*, 1977, *3*, 673-685.

Cooper, F. S., Liberman, A. M., Harris, K. S., & Grubb, P. M. Some input-output relations observed in experiments on the perception of speech. *Proceedings of the Second International Congress of Cybernetics*. Namur, Belgium: Association International de Cybernetique, 1958, 930-941.

Cooper, F. S., & Mattingly, I. G. Computer-controlled PCM system for investigation of dichotic speech perception. *Journal of the Acoustical Society of America*, 1968, *46*, 115. (Abstract)

Cooper, W. E. Selective adaptation to speech. In F. Restle, R. M. Shiffrin, N. J. Castellan, H. R. Lindman, & D. B. Pisoni (Eds.), *Cognitive theory* (Vol. I). Hillsdale, N.J.: Erlbaum, 1975, 23-54.

Cutting, J. E., & Rosner, B. S. Categories and boundaries in speech and music. *Perception & Psychophysics*, 1974, *16*, 564-570.

Cutting, J. E., Rosner, B. S., & Foard, C. F. Perceptual categories for musiclike sounds: Implications for theories of speech perception. *Quarterly Journal of Experimental Psychology*, 1976, *28*, 361-378.

Diehl, R. Feature analyzers for the phonetic dimension *stop* vs *continuant*. *Perception & Psychophysics*, 1976, *19*, 267-272.

Diehl, R. L., Elman, J. L., & McCusker, S. B. Contrast effects on stop consonant identification. *Journal of Experimental Psychology: Human Perception and Performance*, 1978, *4*, 599-609.

Eimas, P. D., & Corbit, J. D. Selective adaptation of linguistic feature detectors. *Cognitive Psychology*, 1973, *4*, 99-109.

Eimas, P. D., & Miller, J. L. Effects of selective adaptation on the perception of speech and visual patterns: Evidence for feature detectors. In R. D. Walk & H. L. Pick, Jr. (Eds.), *Perception and experience*. New York: Plenum Press, 1978, 307-345.

Gerstman, L. J. *Perceptual dimensions for the frication portion of certain speech sounds*. Unpublished doctoral dissertation, New York University, 1957.

Liberman, A. M. Some characteristics of perception in the speech mode. In D. A. Hamburg (Ed.), *Perception and its disorders: Proceedings of the Association for Research on Nervous and Mental Disorders*. Baltimore:

Williams and Wilkins, 1970, 238-254.

Liberman, A. M., Cooper, F. S., Shankweiler, D. P., & Studdert-Kennedy, M. Perception of the speech code. *Psychological Review*, 1967, *74*, 421-461.

Liberman, A. M., Mattingly, I. G., & Turvey, M. T. Language codes and memory codes. In W. A. Melton & E. Martin (Eds.), *Coding processes in human memory*. Washington, D.C.: V. H. Winston, 1972, 307-334.

Lisker, L., & Abramson, A. S. A cross-language study of voicing in initial stops: Acoustical measurements. *Word*, 1964, *20*, 384-422.

Lisker, L., & Abramson, A. S. Distinctive features and laryngeal control. *Language*, 1971, *47*, 767-785.

Locke, S., & Kellar, L. Categorical perception in a nonlinguistic mode. *Cortex*, 1973, *9*, 355-369.

MacNeilage, P. F. Motor control of serial ordering of speech. *Psychological Review*, 1970, *77*, 182-196.

Miller, J. D., Weir, C. C., Pastore, R. E., Kelly, W. J., & Dooling, R. J. Discrimination and labelling of noise-buzz sequences with varying noise-lead times: An example of categorical perception. *Journal of the Acoustical Society of America*, 1976, *60*, 410-417.

Pisoni, D. B. Identification and discrimination of the relative onset of two component tones: Implications for voicing perception in stops. *Journal of the Acoustical Society of America*, 1977, *61*, 1352-1361.

Remez, R. E. *An hypothesis of event-sensitivity in the perception of speech and bass violins*. Unpublished doctoral dissertation, University of Connecticut, 1978.

Remez, R. E. Adaptation of the category boundary between speech and nonspeech: A case against feature detectors. *Cognitive Psychology*, 1979, *11*, 38-57.

Samuel, A. G., & Newport, E. L. Adaptation of speech by nonspeech: Evidence for complex acoustic cue detectors. *Journal of Experimental Psychology: Human Perception and Performance*, 1979, *5*, 563-578.

Schelleng, J. C. The bowed string and the player. *Journal of the Acoustical Society of America*, 1973, *53*, 26-41.

Simon, H. J., & Studdert-Kennedy, M. Selective anchoring and adaptation of nonphonetic continua. *Journal of the Acoustical Society of America*, 1978, *64*, 1338-1357.

Stevens, K. N. The quantal nature of speech: Evidence from articulatory-acoustic data. In E. E. David & P. B. Denes (Eds.), *Human communication: A unified view*. New York: McGraw-Hill, 1972, 51-66.

Weisstein, N. What the frog's eye tells the human brain: Single cell analyzers in the human visual system. *Psychological Bulletin*, 1969, *72*, 157-176.

PRESERVATION OF VOCAL TRACT LENGTH IN SPEECH: A NEGATIVE FINDING*

Betty Tuller+ and Hollis L. Fitch+

Abstract. A primary determinant of vowel quality is vocal tract shape, one aspect of which is vocal tract length. It has been suggested (Perkell, 1969; Riordan, 1977) that vocal tract length is controlled directly, and that one mechanism for its regulation is a coordination between labial and laryngeal gestures. Riordan (1977) observed compensatory changes in the vertical position of the larynx when the characteristic lip protrusion of a rounded vowel was impeded. Although subjects in this study accurately produced the vowels /i/, /a/, /u/ and /ʌ/ with different amounts of lip protrusion, no compensatory larynx height adjustments were observed.

When the movement of an articulator is restricted, speakers are able to produce perceptually acceptable vowels (e.g., Lindblom & Sundberg, 1971; Lindblom, Lubker, & Gay, 1979; Lindblom, Lubker, & McAllister, 1977). It may be that restricted movement of an articulator is accompanied by compensatory vocal tract adjustments. These adjustments would maintain vocal tract shape within the set of physiologically possible configurations that result in equivalent acoustic outputs (Fant, 1960; Nooteboom, 1970; Ladefoged, DeClerk, Lindau, & Papçun, 1972).

One aspect of vocal tract shape that compensatory articulations may preserve is vocal tract length. Indeed, Riordan (1977) has reported that speakers of French or Mandarin Chinese show compensatory lowering of the larynx when lip protrusion is restricted mechanically during the production of front rounded vowels. The vowels produced without normal lip rounding were acoustically similar to the normally produced vowels, indicating the preservation of an acoustic or perceptual target over different vocal tract configurations. The experiment reported here was an attempt to substantiate this finding and to extend it to four vowels (rounded and unrounded) when different amounts of lip protrusion were voluntarily produced by a speaker.

---

*Accepted for publication by the Journal of the Acoustical Society of America (in press).

+Also at University of Connecticut, Storrs.

Acknowledgment: This work was supported by NINCDS grants NS-13617, NS-13870 and BRS grant RR-05596 to Haskins Laboratories. We wish to thank W. Ewan, T. Gay and S. Josell for making equipment and facilities available to us and ensuring that all ran smoothly, and K. S. Harris and J. A. S. Kelso for comments on an earlier draft.

METHOD

Subjects were asked to produce four vowels (/i,a,u,ʌ,/) at each of three lip positions: protruded, flat (relaxed against the teeth) and some position intermediate between the two. For each utterance, movements of the upper lip and of the larynx were monitored photoelectrically and simultaneous acoustic recordings were made.

Subjects

The three adult males chosen as subjects each had a visibly protruding thyroid prominence. This criterion was used in order to facilitate the photoelectric monitoring. Two of the subjects were native speakers of American English and one was a native speaker of British English. All were naive to the purpose of the experiment, and were paid for their voluntary participation.

Procedure

One week before the experiment, subjects were asked to practice the three lip positions until they could consistently produce three distinct amounts of lip protrusion. A stimulus list was read by the subjects during the experiment. The list indicated the vowel and amount of lip protrusion to be produced. Each vowel occurred three times in succession--once at each lip position. The trials were thus blocked in order to minimize the effect of any drift in calibration of the recording apparatus or changes in a subject's head position. All possible orders of the three lip positions occurred for each vowel. The order of the vowels was rotated throughout the list. Twelve tokens of each vowel at each lip position occurred in all.

Lip protrusion and vertical larynx height were monitored photoelectrically, using an improved version of the thyroumbrometer (Ewan & Krones, 1974). A dc light source cast the shadow of the subject's upper lip and thyroid prominence onto separate arrays of photocells. Positions of the lip and larynx were computed from the photocell voltages by a PDP 11/34 computer. The computer output voltage was a staircase function, each step change indicating a .5 mm change in articulator position. Simultaneous acoustic recordings were made so that on subsequent analysis, the acoustic signal and signals corresponding to movements of the lip and larynx could be aligned accurately. The first visible pitch pulse was the point chosen for aligning all tokens of a vowel uttered with the same lip position, and is represented by the zero point on the abscissa in Figure 1. All tokens were judged by the experimenters to be acceptable instances of the intended vowel and none were excluded from the analysis. Moreover, a subset of five tokens of each vowel was checked at a point 200 msec after vowel onset to determine whether the formant frequencies of a given vowel produced with differing amounts of lip protrusion were comparable. Formants were measured by hand from computer-generated spectral displays. Tokens of each vowel at each lip position were averaged using the Haskins Laboratories EMG processing system (Kewley-Port, 1973, 1974). Upper lip protrusion and vertical larynx position were measured 200 msec after the acoustic reference point, near the mid-point of the acoustic waveform.

RESULTS

For all vowels, mean upper lip protrusion in the 'protruded' position was significantly greater than mean upper lip protrusion in the 'flat' position ($p<.001$, for all subjects). Attainment of the mid-range position was not consistent. This indicates that all subjects were able to follow the instructions to the extent of producing two distinct lip positions.

Table 1 presents the mean formant frequencies and standard deviations for each vowel produced with the lips flat and with the lips protruded. Each mean represents formant measurements of five tokens. A series of t-tests revealed no significant differences in the formant values between conditions.

For all three subjects' productions of all vowels, t-tests revealed no significant differences in mean larynx position between the flat and protruded lip conditions. To compare with Riordan (1977), who observed compensatory vertical larynx displacements on a speaker's first attempt to produce the rounded vowel /u/ without protruding the lips, we present in Figure 1 the mean lip and larynx positions for each subject's productions of /u/. None of the differences in larynx height are significant.

DISCUSSION

These results support Riordan's observation that speakers do not raise their larynx when their lips are abnormally protruded during the production of a normally unrounded vowel. In contrast with Riordan's study, however, speakers who were not allowed to protrude their lips for production of rounded vowels did not demonstrate compensatory vertical larynx displacements. Differences in the results observed by Riordan and those reported here may stem from the different methods used. Riordan's subjects were "mechanically restrained" from lip protrusion, whereas these subjects deliberately attempted to produce different lip positions. Riordan obtained vowel samples in CVC syllables, which in turn were embedded in a carrier sentence. In the study reported here, speakers produced isolated vowels with list intonation. Moreover, Riordan's subjects were speakers of French or Mandarin Chinese, whereas these subjects were speakers of English. Nevertheless, no compensatory laryngeal movements were observed in the three speakers who participated in this study, although formant patterns generally were preserved.

Either protruding the lips or lowering the larynx in the absence of other vocal tract changes, tends to lower all formant frequencies. Using one movement to compensate for the absence of the other may not be a generally useful strategy, however, because as Riordan points out, lowering the larynx may change the shape of the vocal tract as well as lengthen it (Lindblom & Sundberg, 1971; Sundberg, 1968). Furthermore, lingual or pharyngeal adjustments may alter vocal tract shape in such a way as to lower formant frequencies. Whatever articulatory compensations may have occurred to preserve vowel identity, such as lingual or pharyngeal adjustments, they did not include changes in laryngeal height. For these speakers, it was not necessary to preserve "vocal tract length" in order to preserve vowel identity.

Table 1

Mean formant frequencies and standard deviations (in parentheses) for three subjects. Values represent five tokens each of the vowels [i,a,u], produced with the lips flat and with the lips protruded.

Formant Frequency (Hz)

		Lips Flat				Lips Protruded			
		F_1		F_2		F_1		F_2	
Subject	Vowel	Mean	sd	Mean	sd	Mean	sd	Mean	sd
BB	/i/	252	(26.1)	2186	(66.2)	266	(32.1)	2184	(67.3)
	/a/	698	(45.5)	1088	(66.1)	716	(23.0)	1068	(42.1)
	/u/	344	(32.1)	974	(131.8)	338	(19.2)	856	(136.7)
BW	/i/	278	(26.8)	1990	(17.3)	290	(29.2)	2000	(18.7)
	/a/	726	(27.9)	1138	(27.8)	706	(11.4)	1120	(31.6)
	/u/	313	(23.9)	970	(120.8)	318	(23.9)	930	(99.5)
TJ	/i/	224	(32.9)	2164	(41.6)	252	(25.9)	2172	(58.1)
	/a/	694	(8.9)	998	(11.0)	696	(8.9)	998	(8.4)
	/u/	282	(19.2)	896	(18.2)	302	(31.9)	888	(25.9)

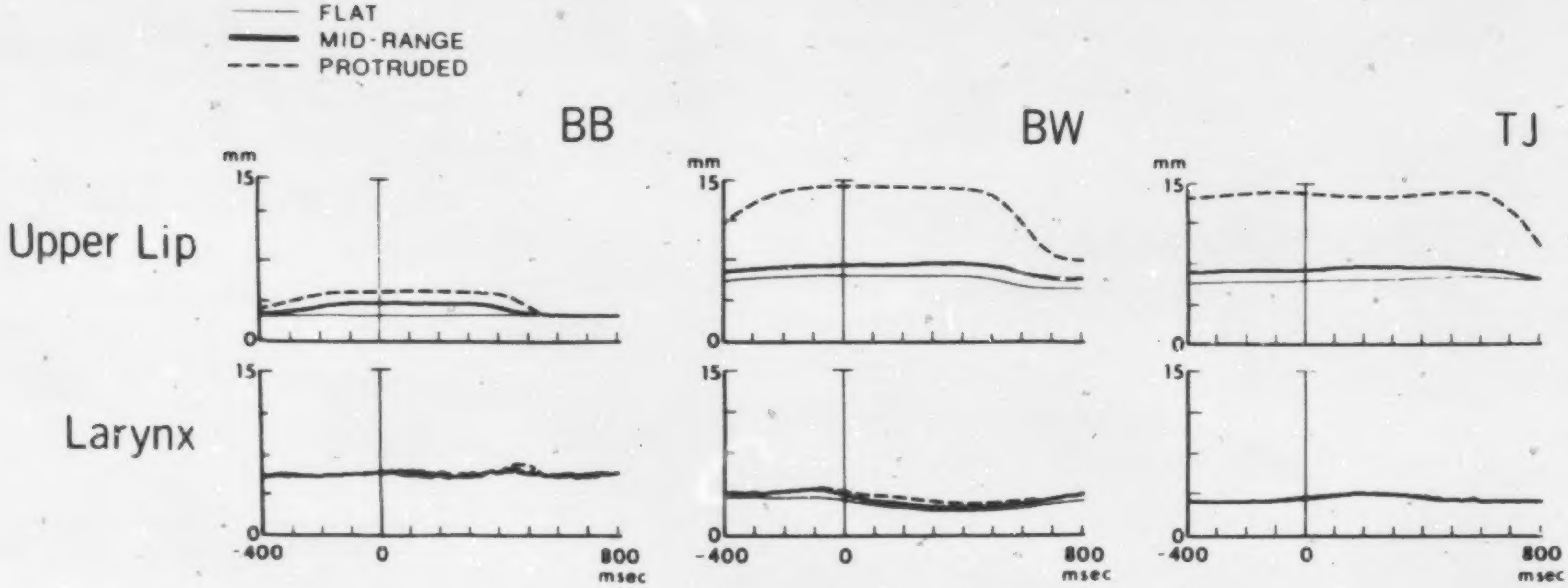


Figure 1. Mean lip and larynx positions, in mm, for each subject's productions of /u/. The zero point on the ordinate was chosen arbitrarily.

REFERENCES

Ewan, W., & Krones, R. Measuring larynx movement using the thyroumbrometer. *Journal of Phonetics*, 1974, *2*, 327-335.

Fant, G. *Acoustic theory of speech production*. 's Gravenhage: Mouton, 1960.

Kewley-Port, D. Computer processing of EMG signals at Haskins Laboratories. *Haskins Laboratories Status Report on Speech Research*, 1973, *SR-33*, 173-183.

Kewley-Port, D. An experimental evaluation of the EMG data processing system: Time constant choice for digital integration. *Haskins Laboratories Status Report on Speech Research*, 1974, *SR-37/38*, 65-72.

Ladefoged, P., DeClerk, J., Lindau, M., & Papçun, G. An auditory-motor theory of speech production. *UCLA Working Papers in Phonetics*, 1972, *22*, 48-75.

Lindblom, B., Lubker, J., & Gay, T. Formant frequencies of some fixed-mandible vowels and a model of speech motor programming by predictive simulation. *Journal of Phonetics*, 1979, *7*, 147-161.

Lindblom, B., Lubker, J., & McAllister, R. Compensatory articulation and the modeling of normal speech production behavior. In R. Carré, R. Descout, & M. Wajskop (Eds.), *Articulatory modeling and phonetics*. Grenoble: G.A.L.F. Groupe de la Communication Parlee, 1977, 147-162.

Lindblom, B., & Sundberg, J. Acoustical consequences of lip, tongue, jaw and larynx movement. *Journal of the Acoustical Society of America*, 1971, *50*, 1166-1179.

Nooteboom, S. G. The target theory of speech production. *IPO Annual Progress Report*, 1970, *5*, 51-55.

Perkell, J. *Physiology of speech production*. Cambridge, Mass.: MIT Press, 1969.

Riordan, C. J. Control of vocal-tract length in speech. *Journal of the Acoustical Society of America*, 1977, *62*, 998-1002.

Sundberg, J. Formant frequencies of bass singers. *Quarterly Progress and Status Report* (Speech Transmission Laboratory, Royal Institute of Technology, Stockholm), 1968, *1*, 1-16.

EPIMENIDES AT THE COMPUTER*

Ignatius G. Mattingly+

If "computer science" is indeed a science, it is in part because the languages in which the programmer communicates with the computer are akin to axiomatized formal systems, such as the propositional calculus or the number-theory system of Russell and Whitehead's *Principia Mathematica*. Even though a set of statements in such a formal system is ordinarily a proof, while a set of statements in a computer language is ordinarily a routine that instructs the computer to execute a sequence of logical steps, certain central concepts of mathematical logic have been shown to be directly relevant to computer programming.

One such concept is that of recursive definition. A recursive function, for example, the definition of a series of numbers $G(0)$, $G(1)$ ... $G(n)$ ...:

$$G(n) = n - G(G(n-1)) \text{ for } n > 0$$
$$G(0) = 0$$

cannot be evaluated for an arbitrary value of n by conventional algebraic techniques because one cannot immediately compute $G(n-1)$, let alone $G(G(n-1))$. But, knowing that $G(0) = 0$, one can determine that for $n = 1$, $G(n-1) = 0$, so $G(G(n-1) = (0)$, and so $G(n) = 1$. Knowing that $G(n) = 1$ for $n = 1$, one can, by performing a series of iterative calculations, each depending on the result of its predecessor, evaluate $G(n)$ for $n = 2, 3, \ldots$ until the required value of n is reached. Such calculations, tedious and error-prone when carried out by a human being, are just what computers are good at, and an experienced programmer will try to cast the problem he wishes to solve in the form of a recursive definition.

A closely related notion is that of nested logical structure. An extremely complex proof in the propositional calculus can be made perspicuous if it can be organized as a group of subordinate derivations, and a subordinate derivation may in turn have subordinate derivations of its own, the process extending to whatever depth of nesting is required (obviously, if the form of each successively nested derivation is the same, the proof is simply recursive). Analogously, a computer program can be organized as a series of calls to subroutines, each of which may in turn call other subroutines, and so on.

---

*A review of *Gödel, Escher, Bach; An Eternal Golden Braid*, by Douglas R. Hofstadter (New York: Basic Books, 1979), *Yale Review*, Winter, 1980, 270-276.

+Also University of Connecticut, Storrs.

Acknowledgment: Support from NICHD Grant HD01994, National Institutes of Health, is gratefully acknowledged.

[HASKINS LABORATORIES: Status Report on Speech Research SR-59/60 (1979)]

Again, logicians have long been interested in the logical status of self-referent sentences because they can be paradoxical (e.g., Epimenides the Cretan's assertion that "all Cretans are liars," or in other words, "this statement is false"). Not all self-referent statements in natural language lead to paradox, however (e.g., "This statement is in English") and all computer languages of practical interest allow self-reference, because this property permits a program to modify itself as the computation proceeds. Part of the routine for executing the nth step of the evaluation of a recursive function is a slight modification that converts it into a routine for executing the n+1st step. It is the property of self-reference in the programming language that essentially distinguishes a computer from a non-programmable calculator.

The programmer, however, must be very careful with recursive procedures, nested structures and self-reference. If the number of iterations or the initial conditions of a recursive routine or the calls to nested subroutines are improperly specified, if they are self-referent in a way he does not intend, his program may put the computer into an infinite regress that can be halted only by human intervention. It would be a great help, therefore, if only a general "checking program" could be devised that could inspect any program in a given computer language and determine whether it will always terminate. Unfortunately, such a checking program can be shown to be not just impractical but impossible, and the explanation for the impossibility is to be found in still another insight of mathematical logic, Gödel's theorem, which says that for any axiomatic system that permits self-reference, there will be well-formed, true statements that cannot be derived from the axioms (e.g., Epimenidean self-referent statements that can be paraphrased, "This statement is not a theorem of the system").

Douglas R. Hofstadter is a computer scientist who has clearly thought a great deal about these rather difficult mathematical and logical ideas. For him, they are the key to the understanding not just of computer programs, but of language, art, and the mind itself, and he is anxious to communicate this view to a general audience. "In a way," he says, "this book is a statement of my religion. I hope that this will come through to my readers and that my enthusiasm and reverence for certain ideas will infiltrate the hearts and minds of a few people" [p. xxi].

In order to explain these ideas and the applications he wishes to make, he has adopted a rather unorthodox pedagogical strategy. Instead of simply developing his argument step by step, he shifts back and forth from one theme to another, expounding the same idea repeatedly at increasing levels of complexity. This organization is, however, not haphazard; it is a deliberate attempt to parallel musical form, in particular the form of Bach's A Musical Offering, the fugues and canons in which are supposed to illustrate the very formal structures in which Hofstadter is interested. The logical ideas are presented in a number of expository chapters that include extensive analogies from music and graphic art (this is where Bach and Escher come in), as well as from the natural sciences and Zen Buddhism. The style is clear but hardly elegant ("Gödel realized that there was more here than meets the eye" [p. 18]). Interspersed with the expository chapters are a number of whimsical dialogues, supposed to be parallel in form to various compositions of Bach, in which Achilles hashes over these same ideas with various talking animals, after the manner of Lewis Carroll (one of the dialogues is in fact a reprint of Carroll's "What the Tortoise said to Achilles").

Hofstadter's pedagogical strategy has the undoubted advantage that even the most abstruse concepts begin to sink in on the _n_th iteration. Its disadvantage is that, as each new topic is introduced, the reader must have faith that it will eventually prove to be relevant to the main argument. Unfortunately, Hofstadter does many things to weaken one's faith. He is given to quoting long passages needlessly. He includes a great deal of extraneous historical detail (much of it fascinating enough if one hasn't heard it before) about Bach, Babbage, Gödel, Turing, Fermat, Cantor, and others. He decorates the book with portraits of these figures (including one of Turing in athletic costume), as well as hundreds of other unnecessary illustrations. He continually offers such pointless observations as "It is interesting to note that the lives of Mumon and Fibonacci coincided almost exactly: Mumon living from 1183 to 1260 in China, Fibonacci from 1180 to 1250 in Italy" [p. 246]. He cracks a lot of rather poor jokes (one of the dialogues is entitled "SHRDLU toy of man's designing") and then makes matters worse by explaining them. This kind of thing obscures the argument, mars the structure of the book, and makes it much longer (777 pages) than necessary.

In spite of these excesses, it must be said that Hofstadter's "enthusiasm and reverence" for the ideas that fascinate him certainly do come through on every page. And as long as he is explaining the ideas themselves, or relating them to computer programming, he is powerful, lucid, and persuasive. The reader who has found conventional textbook presentations of the ideas of Gödel and Turing difficult to penetrate may well be beguiled into understanding by Hofstadter, and even the reader who already has some glimmerings may find his appreciation of these ideas considerably deepened.

But when Hofstadter tries to demonstrate their pervasiveness in other areas he is not so convincing. As his title suggests, he feels that Bach and Escher, because of their use of recursive structure and self-reference, have some affinity with Gödel. The relevance of these concepts to Escher's drawings is obvious enough: in "Ascending and Descending," for example, a column of monks plods up a stairway whose top somehow appears to be its bottom, and in "Drawing Hands," two hands appear to be drawing each other (these two drawings by Escher, and no less than 32 others, are included among the illustrations). The argument with respect to Bach seems rather less cogent. Though Bach has many ways of varying a theme, and frequently modulates from key to key in a systematic pattern, Hofstadter can really offer only two convincing examples of nested or recursive structure: the "Canon per tonos" in _A Musical Offering_ and the _Little Harmonic Labyrinth_. The example of "self-reference": the use of the notes B, A, C, H in the incomplete _Art of the Fugue_, is foolish and irrelevant, and leads only to the bizarre suggestion that Bach fell ill and died before completing this work because of his "attainment of self-reference" [p. 86]. Hofstadter's fascination with Bach's structural devices is as manifest as his fascination with the properties of formal systems, but his "braiding" of the two together often leaves one confused rather than convinced.

Hofstadter is on much firmer ground when he considers the structure of human language. A strong case can be indeed made for including recursive rules in grammars. As Noam Chomsky has argued, only in this way can one account for the ability of a finite grammar to generate an infinite number of sentences. However, a theory of grammar that simply allowed the unrestricted use of recursive devices would be too powerful: It would permit not only grammars that can occur in natural languages but also an infinite number that cannot. This is the objection to the theory of grammar implicit in an "Augmented Transition Network," a type of recursive procedure which has been used with considerable success by Terry Winograd and others in computer programs for parsing English sentences, and which Hofstadter takes seriously as a model of human sentence parsing. The real problem of the linguistic theoretician is to constrain a grammatical theory permitting recursive devices so that it permits just those grammars that can occur. Hofstadter does not appreciate this point, perhaps because he is, it would appear, aware of current linguistic theory only at second hand: He does not even mention Chomsky.

Having made forays into crystallography and nuclear biology, Hofstadter turns to the problem of modeling human intelligence. Mathematics can not be done except by computation, he argues; since a human mind can solve mathematical problems, its machinery must include some general recursive function for sorting numbers into two classes (this is the "Church-Turing thesis"). What is true of this presumably clear case of human intelligence in action must also be true of other, less well-defined cases. If so, given a non-trivial computer language, it should be possible to write computer programs that simulate other mental activities, and these programs, if successful, must be viewed as veridical models. Such programs, in fact, form the agenda of those computer scientists (of whom Hofstadter is one), who are practitioners of "artificial intelligence." As is well-known, programs have been written that, with varying degrees of success, play chess and checkers, recognize visual and acoustic patterns, synthesize speech, and parse sentences. Eventually, it is suggested, the human mind will be modeled as one large but coherent computer program.

The objection to this argument is not that the Church-Turing thesis is false, but that the extremely modest nature of the psychological claim it makes is disguised. There are uncountable different ways, all compatible with the Church-Turing thesis, in which a human being might conceivably go about solving any particular class of problems, so that a program that models one of these ways is not necessarily of any psychological interest. The mere fact that the program successfully solves the problems set for it, though it may be an impressive demonstration of the programmer's ingenuity, is far from being psychologically conclusive. Indeed, the remarkable success of Arthur Samuels' checker-playing program arouses the suspicion that the specific strategies it uses are quite different from those used by a human player. If so, the program may be telling one a great deal about checkers but not very much about the human mind. Whether the problem is checker-playing or sentence-parsing, the objective should be the development, not of a merely successful program, but of a program that is constrained by what is known of the strategies, effective or not so effective, that human beings actually use. As much recent research in psycholinguistics demonstrates, these strategies can indeed be studied and described.

Is Hofstadter really saying anything save that science is logical? A rigorous model of any natural process must in principle be expressible in a formal system. As he himself makes clear, all but the most trivial of formal systems must allow self-reference and recursive devices, and hence must be subject to the logical limitations expressed by Gödel's theorem. If the model is to make any interesting empirical claims, therefore, it must propose additional constraints of some kind; it is the precise character of these constraints, as has been insisted, that is of primary importance to the physicist, the biologist, or the psychologist. In the absence of such proposals, Hofstadter's arguments come close to being vacuous.

LANGUAGE BY HAND AND BY EYE*

Michael Studdert-Kennedy+

Language is form, not substance. Yet every semiotic system is surely constrained by its mode of expression. Communication by odor, for example, is limited by the relatively slow rates at which volatile chemicals disperse and smell receptors adapt. By the same token, we might suppose that the nature of sound, temporally distributed and rapidly fading, has shaped the structure of language. But it is not obvious how. What properties of language reflect its expressive mode? What properties reflect general cognitive constraints necessary to any imaginable expression of human language? How far are those constraints themselves a function of the mode in which language has evolved?

Until recently, such questions would hardly have been addressed, because we had no unequivocal example of language in another mode, and because there are grounds for believing that language and speech form a tight anatomical and physiological nexus. Specialized structures and functions have evolved to meet the needs of spoken communication: vocal tract morphology, lip, jaw and tongue innervation, mechanisms of breath control, and perhaps even matching perceptual mechanisms (Lenneberg, 1967; Lieberman, 1972; Du Brul, 1977). Moreover, language processes are controlled by the left cerebral hemisphere in over 95% of the population, and this lateralization is correlated with left-side enlargement of the posterior planum temporale (Geschwind & Levitsky, 1968), a portion of Wernicke's area, adjacent to the primary auditory area of the cortex and known to be involved in language representation. Wernicke's area is itself linked to Broca's area, a portion of the frontal lobes, adjacent to the area of the motor cortex that controls muscles important for speech, including those of the pharynx, tongue, jaws, lips, and face; damage to Broca's area may cause loss of the ability to speak grammatically, or even to speak at all. Taken together, such facts suggest that humans have evolved anatomical structures and physiological mechanisms adapted for communication by speech and hearing.

Furthermore, the structure of spoken language, based on the sequencing of segments, follows naturally from its use of sound, that is, of rapid variations in pressure distributed over time. At the level of syntax, the segments are words and other morphemes. At the level of the lexicon, the segments are phonemes (consonants and vowels) arranged in sequences to form syllables and words. This dual pattern of sound and syntax, commonly cited as a distinctive property of language, perhaps evolved to circumvent limits on our capacity to produce and perceive sounds. Certainly, the number of

---

*A review of The Signs of Language by Edward S. Klima and Ursula Bellugi (Cambridge, Mass.: Harvard University Press, 1979). This review will be published in Cognition.

+Also Queens College and Graduate Center, City University of New York.

Acknowledgment: Preparation of this review was supported in part by Grant HD-01994 from the National Institute of Child Health and Human Development.

holistically distinct sounds that the human vocal apparatus can make and the human ear perceive, is relatively small. Perhaps in consequence all spoken languages construct their often vast lexicons from a few (usually between about 20 and 60) arbitrary and meaningless sounds, and set restrictions on the sequences in which the sounds may be combined.

The sounds selected and the rules for their combination differ from language to language, but all languages make a major class division between consonants, formed with a more-or-less constricted vocal tract, and vowels, formed with a relatively open tract. The division reflects a natural opposition between opening and closing the mouth, and is therefore peculiar to speech. The combination of consonant and vowel gestures into a single ballistic movement gives rise to the consonant-vowel syllable, a fundamental articulatory and acoustic unit of all spoken languages. The acoustic structure of the syllable departs from the rule of sequence, since parallel or co-articulation of consonant and vowel yields an integral event in which acoustic cues to the two components are interleaved. However, this departure may itself be an adaptation to limits on hearing, short-term memory and the cognitive processes necessary to understand a spoken utterance. If we hypothesize an ideal speaking rate--neither too slow nor too fast for comfortable comprehension--and take, as a measure of this ideal, a standard English rate of about 150 words a minute, the phoneme rate (allowing, say, 4 phonemes per word) will be 10 per second, close to the threshold at which discrete acoustic events merge into a buzz. By packaging consonants and vowels into the basic rhythmic unit of the syllable, speech reduces the segment rate to a level within the temporal resolving power of the ear (Liberman, Cooper, Shankweiler, & Studdert-Kennedy, 1967).

In short, the dual pattern of lexical form and syntax, the detailed acoustic structure by which lexical form is expressed, and what little we know of the neurophysiology of speech and language, all suggest that speech is the natural, and perhaps even necessary, mode of language. But the advent of systematic research into sign languages, employing a manual-visual spatial mode rather than an oral-auditory temporal mode, has made it possible to test this assumption and to ask fundamental questions about language and its organization. Can language be instantiated in another mode? If so, how is it organized? Does it display a dual structure of lexical form and syntax? How are its formational and grammatical functions realized within the constraints of hand and eye rather than of mouth and ear?

Sign languages are of two types (Stokoe, 1974). The first type is artificial and is based, like writing and reading, on a specific spoken language: Its signs refer to letters ("fingerspelling") or higher-order linguistic units (words, morphemes), and its syntax follows that of the base language. Examples are the sign languages of Trappist monasteries, of industrial settings, such as sawmills, and the various sign languages of the deaf (e.g., Signed English), developed and largely used in schools to facilitate reading and writing. The second type is not an artifact: it is not based on any spoken language. Rather, both lexicon and syntax are independent of the language of the surrounding community or of any other spoken language. Examples are the sign languages of the Australian aborigines, of the American Plains Indians (West, 1960; Umiker-Sebeok & Sebeok, 1977) and of deaf communities all over the world. An important distinction is drawn by Stokoe (1974) between aboriginal and deaf sign languages. The former are usually learned as a second language by individuals who already know a spoken language. The latter are usually learned as a first language by congenitally deaf infants, and are ontogenetically free from contamination by spoken language. The most extensively studied deaf language has been American Sign Language (ASL), said by Mayberry (1978) to be the fourth most common language in the United States.

Modern ASL derives from a French-based sign language, codified by the Abbe de L'Epée in the 18th century and introduced to the United States by Thomas Gallaudet in 1817. (Users of ASL today find French SL more intelligible than British SL [Stokoe, 1974]--evidence for the independence of ASL from the surrounding language.) Early French sign language, and its American counterpart, were combinations of lexical signs originating among the deaf themselves and of grammatical signs corresponding to French (or English) formatives introduced by de L'Epée and his followers to help deaf pupils to learn to read and write. However, these speech-based signs rapidly fell into disuse--presumably because they ran up against the natural tendency of sign languages to conflate rather than concatenate their morphemes--and for the past 160 years ASL has developed among the deaf as an independent language (although see Fischer, 1978, for a discussion of ASL as an English-based creole).

Until recently, established wisdom regarded sign languages of the deaf, like that of the Plains Indians, as more-or-less impoverished hybrids of conventional iconic gesture and impromptu pantomime. Analysis of their internal structure was limited to description of the images suggested by the forms of signs.[2] The first steps toward a structural description of ASL were taken by Stokoe (1960). With the publication of A Dictionary of American Sign Language on Linguistic Principles (Stokoe, Casterline, & Croneberg, 1965), containing an account of nearly 2500 signs, the study of ASL entered a new period. Stokoe and his colleagues showed that signs were differentiated along three dimensions, or parameters: handshape, place of articulation, and movement. On the basis of a minimal pair analysis, they posited a limited set of distinctive values, or primes, on these dimensions: 19 for handshape, 12 for place of articulation and 24 for movement, making a total of 55 "cheremes," analogous to the phonemes of a spoken language. By demonstrating the existence of sublexical structure, Stokoe opened the way for systematic research into ASL and its relation to spoken language.

The task was undertaken by Edward Klima and Ursula Bellugi, and has been the focus of an ambitious program of research for the past seven years at the Salk Institute for Biological Studies in La Jolla, California. The present book is a brilliant recension of that research, extending Stokoe's original analysis, supplementing it with an imaginative range of linguistic and psycholinguistic studies and, for the first time, revealing some of the complex grammatical processes by which ASL combines and elaborates its lexical units.

The authors strictly observe the distinction between linguistic and psycholinguistic analysis. The book is divided into four parts. Part I undertakes to separate iconic invention from arbitrary structure; Part II reports a series of psycholinguistic studies of short-term memory, slips of the hand, and the featural properties of signs; Part III returns to linguistic analysis with an extended investigation of grammatical processes; Part IV concludes the book with an account of wit, play, and poetry. The subject matter may seem difficult, even forbidding, to the glottocentric reader, like myself, who knows no sign language and is taxed by the effort of imagining the complex, three-dimensional shapes and movements by which ASL conveys its messages. But the exposition is simple, precise, and so richly illustrated with photographs and detailed drawings (roughly one every three pages) that one soon forgets one's ignorance and is absorbed in the argument of the text. The work, marked throughout by analytic rigor, depth, and weight, is unquestionably the most thorough and detailed study to date of any sign language.

The focus of the book is on the effects of modality. Its aim is to broaden and deepen understanding of language by sifting finer properties peculiar to language mode from more general properties common to all forms of linguistic expression. The most pervasive property of ASL (and, doubtless, of every manual sign language) is its iconicity. Signs are often global images of some aspect of their referents, their grammar is often marked by congruence between form and meaning, and casual discourse grades easily into gesture and mime. Such mimetic processes are themselves worthy of study (e.g., Friedman, 1977), for they certainly reflect human cognitive and semiotic capacity--what other animal is capable of the "excellent, dumb discourse" of pantomime? But ASL is also abstract, and the first task for the analyst is to separate what the authors call "the two faces of sign: iconic and abstract."

The iconic itself has two faces: first, the extrasystemic pantomime that may accompany signing; second, the iconic properties of the lexical signs themselves. Of course, a modest pantomime often accompanies speech--imagine an excited account of a car crash--but we have no difficulty in separating vocal from bodily gesture because the two types follow different channels of communication. To separate the channels in a sign language is a more delicate task, and one that has defeated many earlier analysts. The authors, with typical directness and ingenuity, solved the problem by asking a deaf mime artist to render a variety of messages in both ASL and pantomime, and to maintain as much similarity between the two renditions as possible. From slow motion playback of his performance they established criteria for separating pantomime from sign. In general, the signed rendition was shorter than the mime (by a factor of 10 to 1), the signs themselves discrete rather than continuous (cf. West, 1960, p. 5), relatively reduced, compressed, and conventionalized. Moreover, in pantomime, the eyes were free to participate in the action, anticipating or following movement of the hands, while, in signing, they made direct contact with the addressee throughout the sign. Thus, by requiring sustained eye contact during signing, ASL limits the visual field within which signals may be made. The perceptual structure of this field for the addressee (fine at its foveal center, coarse at its periphery) then constrains the form and location of signs (Siple, 1978).

Before commenting on the iconic properties of the signs themselves, we should note their range of reference. Some signs translate into a single English word, some into several; others, such as distinct pronominal signs for persons, vehicles, and inanimate objects, have no English counterparts at all. In short, there are thousands of lexical signs in ASL, covering a full range of categories and levels of abstraction. Yet many signs do have obvious iconic components: The sign for "house" traces the outline of roof and walls; the sign for "tree" is an upright forearm, with spread, waving fingers; the sign for "baby" is one arm crossed in front of the other, while the arms rock. Nonetheless, just as we are often unaware of metaphor until it is pointed out ("He's a sharp operator"), non-signers usually cannot judge the meaning of a sign, but, once informed, may readily offer an account of its iconic origin. The "paradox of iconicity," in the authors' phrase, is, first, that icons are conventional, so that quite different aspects of a referent may be represented by different sign languages (Chinese, Danish, British, American, and so on); second, that icons, despite their "translucent" origin, become so modified by the structural demands of the language that their iconicity is effectively lost. Indeed, as Frishberg shows in her chapter on historical change, comparisons of modern ASL signs with those depicted in manuals and films of seventy years ago show a strong tendency for signs to be condensed, simplified, stylized, moving toward increasingly abstract forms, by a process perhaps analogous to the development of figural representation in, for example, Byzantine painting. Similar observations have been made of Plains Indian Sign Language (e.g., Kroeber, 1958, cited by Umiker-Sebeok and Sebeok, 1977, p. 75). Thus, a main goal of the book's argument is to demonstrate, in compelling detail, how arbitrary form and system subdue mimetic representation.

Here, we need some account of the structure of ASL signs. As already noted, Stokoe (1960) and his colleagues (Stokoe, Casterline, & Croneberg, 1965) first described the sublexical structure of ASL citation forms. Various later analysts have proposed slightly different classifications or numbers of primes and subprimes ("phonetic" variants), but all have followed the principle of Stokoe's analysis. Klima and Bellugi, terming the three parameters of variation Hand Configuration, Place of Articulation, and Movement, propose a number of modifications, most of them needed for the analysis of morphological processes not attempted by Stokoe.

Hand Configuration refers to distinct shapes assumed by the hands, and includes a minor parameter of hand arrangement, specifying the number of hands used to make a sign and their functional relation (about 60% of ASL lexical signs use two hands). Place of Articulation refers to the location within signing space (a rough circle, centered at the hollow of the neck, with a diameter from the top of the head to the waist) at which a sign is made or with reference to which it moves (chin, cheek, brow, torso, and so on). Klima and Bellugi further posit a division of the space in front of the signer's torso into three orthogonal planes (horizontal, frontal, sagittal); these abstract surfaces prove important in the description of inflected forms. Movement, the most complex dimension, includes primes that range from delicate hand-internal movements through small wrist actions to the tracing of lines, arcs or circles through space. But a full description of the movement parameter, sufficient to distinguish between certain lexical signs, between lexical categories (such as noun and verb [Supalla & Newport, 1978]) and, especially, among the multitude of richly varied, inflected forms, requires a description of the dynamic qualities of movements: rate, manner of onset or offset, frequency of repetition, and so on.

Structural analysis of ASL is at its beginning, but the lower level of a dual pattern, analogous to that of spoken language, has already begun to emerge. The number of possible hand configurations, places of articulation, types and qualities of movement must be very large. Yet ASL uses a limited set of formational components, analogous to the limited set of phonemes in a spoken language. Moreover, just as spoken language restricts the sequential combination of phoneme types within a syllable, so ASL restricts the simultaneous combination of spatial values within a sign. Some combinations are doubtless difficult, or impossible, for physical reasons. For example, the Symmetry Constraint, posited by Battison (1974), requires that, if both hands move in forming a sign, their shapes, locations, and movements must be identical. Given the well-known difficulty of coordinating conflicting motor acts of the two hands, this rule may prove common to all sign languages. However, other combinatorial constraints seem to be ruled out for arbitrary, language-specific reasons. As preliminary evidence for this, in the absence of a full linguistic analysis of another sign language, the authors adduce psycholinguistic evidence from a comparison of selected signs in Chinese Sign Language (CSL) and ASL. The study showed that certain combinations of handshape, place of articulation and movement primes used in CSL are unacceptable to native signers of ASL, while other CSL combinations are acceptable, but do not occur in ASL.

Thus, linguistic analysis leads to a view of the ASL sign as a complex, multidimensional structure, conveying its distinctive linguistic information by simultaneous contrasts among components arrayed in space rather than by sequential contrasts arrayed in time. As the authors observe, if this arbitrary sublexical structure exists in a language of which the representational scope is so much richer than that of speech, we may reasonably infer that the formational structure of both languages offers more than mere escape from the limits of articulation. We may suspect, rather, a general cognitive function, perhaps that of facilitating acquisition, recognition, recall, and rapid deployment of a sizeable lexicon (cf. Liberman & Studdert-Kennedy, 1978; Studdert-Kennedy, in press).

In Part II of the book the authors report a variety of psychological studies, designed to "...explore the behavioral validity of the internal organization of ASL signs posited on the basis of linguistic analysis" (p. 87). Several studies--of short-term memory for random lists, of slips of the hand in everyday signing, of sign perception through visual noise--are modeled on similar studies of speech, often cited as evidence for the psychological reality of the coarticulated components of the syllable, and they reach strikingly similar conclusions.

The central question of these studies is: In what form do native signers encode and process the signs of ASL? Do sublexical components enter into the coding process? Unequivocally, they do. For example, when native signers, fluent in reading and writing English, were asked to recall random lists of ASL signs and to write their responses in English words, their errors did not reflect either the phonological structure or the visual form of the written words, nor did they reflect the global iconic properties or the meanings of the signs. Instead, errors reflected the signs' sublexical structure, and the most frequent errors differed from the presented sign on a single parameter. By contrast, the intrusion errors of hearing subjects, asked to recall equivalent lists of English words, reflected the phonological structure of the words--the usual result in such studies (see, for example, Conrad, 1972).

These results hint, incidentally, at an answer to the old question of whether intrusion errors in short-term memory for spoken (or written) words are based on similarities in sound or in articulation. The parallel between signs and words suggests that the effects may be based on a coding process common to both speech and sign. Rather than acoustic for speech, visual for sign, short-term memory codes for both modalities may be either motor (cf. Aldridge, 1978) or abstract and phonological (cf. Campbell & Dodd, in press).

That the motor system codes signs along the posited linguistic dimensions is evidenced by errors in everyday signing. The authors analyzed a corpus of 131 slips of the hand, much as comparable speech errors have been analyzed (e.g., Fromkin, 1971), and with analogous results. As in the speech data, most errors were anticipations and perseverations (rather than complete metatheses) of sublexical units--here, values of the structural parameters--and, typically, the errors gave rise to permissible combinations of parametric values that happened not to be items in the lexicon (ruling out lexical substitution as the source of error). The rarity of inadmissible parametric combinations demonstrates the force of formational constraint. The important conclusion is that everyday signing is not a matter of concatenating globally iconic forms, but is sensitive to the internal structure of the signs.

Moreover, native signers are aware of sign structure, just as speakers are aware of word structure. Wit and play (Part IV) are quite different in the two modalities because, while spoken gesture is confined to the hidden space of a vocal tract and can be revealed only by its acoustic effect, signs are executed in the same physical space as the signers themselves occupy. Accordingly, like figures on a Baroque ceiling whose limbs break from their frame into the real space below, signs readily escape into informal gesture or pantomimic elaboration. Nonetheless, structural play does occur. Punning, it seems, is rare, perhaps because ASL has few homomorphs (virtually every distinction of meaning is signaled by a distinction of form). The characteristic mode of sign play is apparently the "...compression of unexpected meanings into minimal sign forms" (p. 320), often by substituting the hand configuration, place of articulation, or movement of one sign for the corresponding parameter of another, to produce a cross between the two, analogous to Lewis Carroll's portmanteau words (e.g., chuckle + snort = chortle). In "art sign," as the authors term the developing poetic (or, perhaps better, bardic) tradition of the National Theater for the Deaf, artists fulfill the cohesive functions of spoken alliteration, assonance, and rhyme by choosing signs that share hand configuration or place of articulation; effects analogous to melody and rhythm they achieve by enlarging, blending, syncopating sign movements into a spatio-temporal kinetic superstructure. In other words, signers display, in both casual humor and formal art, a knowledge of the internal structure of signs.

Up to this point we have treated the values, or primes, of the major parameters as integral units, analogous to the phonemes of spoken language. Indeed, in their early linguistic analyses, the authors found no evidence for formational (i.e., "phonological") rules defining featural classes among the primes, analogous to those posited for phonemes by current linguistic theory. They therefore undertook to reverse the usual direction of research by looking for psycholinguistic evidence of sub-prime features that might later guide (and be validated by) linguistic analysis. They modeled their study on the well-known work of Miller and Nicely (1955). Miller and Nicely, it will be recalled, attempted to test the perceptual reality of certain traditional articulatory features by measuring the systematic feature-based confusions among English, nonsense-syllable consonants offered for identification in random masking noise. Similarly, the present authors videotaped a set of nonsense-signs, incorporating the 20 primes of Hand Configuration, and offered them to native signers for identification in random visual noise. They gathered their results into confusion matrices and derived, by cluster analysis and multidimensional scaling procedures, a set of 11 features that differentiated the 20 hand configurations. The psychological validity of the proposed feature set was suggested by the outcome of other studies: For example, intrusion errors on the recall of Hand Configuration, in the short-term memory studies described above, tended to be on a single feature.

However, since the perceptual study did not include a control group of hearing subjects, we have no way of knowing whether the derived features reflect an abstract "phonology" or mere psychophysical similarities among Hand Configurations.[3] The latter interpretation is encouraged by the outcome of a subsequent study of Place of Articulation in which hearing controls were used (Poizner & Lane, 1978). Here, although the linguistic knowledge of native signers was reflected both by a response bias in favor of places of articulation that occur more frequently in ASL and by greater overall accuracy than hearing controls, scaling and clustering solutions to the confusion matrices of the two groups were essentially the same. Such an outcome for the Hand Configuration study of the present book would have robbed the derived features of even psycholinguistic validity. But, as the authors explicitly state, their "...preliminary model of suggested features...ultimately must depend for its confirmation on its usefulness for linguistic analysis" (p. 178), and this usefulness has yet to be demonstrated.

In any event, we have seen that ASL signs do display a clear sublexical structure to which native signers are sensitive. Evidently, duality of patterning did not evolve, as we first surmised, merely to circumvent limits on speaking and hearing, but, as suggested above, has a more general linguistic function that must be fulfilled in both spoken and signed languages. Can the same be said of the syllable into which the sublexical units of speech are compressed? Certainly, with few exceptions, hand configuration and place of articulation are maintained throughout the movement of a sign, so that ASL exploits its visuo-spatial mode to achieve the ultimate compression of its sublexical units: simultaneity. However, the degree of compression is so much greater for the sign than for the syllable that we may suspect quite different functions. What we need is a broader comparison between the fundamentally temporal structure of speech and the fundamentally spatial structure of sign.

The authors lead into this comparison with several studies on the rates of speaking and signing. Their first discovery, confirmed by Grosjean (1977), was that the average sign takes roughly twice as long to form as the average word takes to say. Their second discovery was that, if the spontaneously signed version and the spontaneously spoken version of a story are divided into propositions--"defining a proposition as something that can be considered equivalent to an underlying simple sentence" (p. 186)--the mean proposition rates for the two versions are roughly equal. These results suggest, first, that ASL has time-saving devices for expressing grammatical relations among signs spatially rather than temporally; second, more generally, that a single, temporally constrained cognitive process may control the proposition rates of both languages.

The authors identify three main spatial devices by which ASL conflates lexical and grammatical information. First is a device often emphasized in accounts of Plains Indian Sign Language (West, 1960): deixis or indexing. ASL achieves pronominal and anaphoric reference by establishing a locus for each of the actors or subjects under discussion. Later reference is then made simply by directing action signs toward the established locus.

A second device, of the utmost importance in demonstrating recursive, syntactic mechanisms in ASL, is the use of facial expression and bodily gesture to indicate clausal subordination. The authors do not elaborate, since they confine their attention in this book to the formational properties of manual signs. But they cite Liddell (1978), who has shown that a relative clause may be marked in ASL by tilting back the head, raising the eyebrows and tensing the upper lip for the duration of the clause. Other non-manual configurations (including blinks, frowns and nods) may mark the juncture of conditional clauses (Baker & Padden, 1978).

The third incorporative device is the modulation of a sign's meaning by changes in the spatial and temporal properties of its movement. Among the many functions of such changes are to differentiate nouns from verbs, modify adjectival and verbal aspect, and inflect verbs for distinctions within a variety of grammatical categories. These modulations are the topic of chapters in Part III, devoted to morphological processes in ASL.

Part III begins with an account of productive grammatical processes by which new signs enter the language. One fertile process is the stringing together of lexical items to form compounds, analogous to English *breakfast*, *kidnap*, *bluebird*. For example, ASL has combined the signs BLUE[4] and SPOT to form a new sign BLUESPOT, meaning "bruise." In English, such compounds are distinguished from phrases by overall reduced duration and by a shift in stress from the second word to the first: *hard hát* (a hat that is hard) becomes *hárdhat* (a construction worker). Similarly, in ASL overall duration is reduced, so that the compound lasts about half as long as the original two signs together, but (the opposite of the English process) reduction of the first sign is roughly twice as great as that of the second. Typically, the first sign reduces its movement, suggesting an incipient blend into a single sign (cf. English: *anise seed* becomes *aniseed*). Even before the blend is complete, the contributing signs will have lost their original meaning. BLUESPOT can refer to a bruise that is yellow, just as *hárdhat* designates a person, not a hat. Similar compounding processes are used in ASL to derive from signs for objects (chair), signs for superordinate (furniture) and subordinate (kitchen chair) lexical categories. The discovery of such grammatical mechanisms for creating new signs (fully analogous to those of many spoken languages) challenges the common notion that sign language lexicons are intrinsically limited and can be expanded only by iconic invention.

But the real breakthrough in morphological analysis was the discovery of changes in the temporal-spatial contours of signs to modify their meaning.

The key insight was that, in its grammar no less than in its lexicon, ASL uses simultaneous rather than sequential variation.5 Modulations of the meaning of a lexical item are achieved not by adding morphemes, as is typical of many spoken languages, but by modifying properties of one of the sign's parameters, its movement. In English, changes in aspectual meaning (that is, distinctions marking the internal temporal consistency of a state or event, such as its onset, duration, frequency, recurrence, permanence, intensity) are made by concatenating morphemes. A single adjectival predicate is used in a range of syntactic constructions to yield different meanings: *he is sick*, *he became sick*, *he gets sick easily*, *he used to be sick*, and so on. In ASL precisely the same modulations of meaning are achieved by changes in the movement of the predicate SICK itself: hand configuration and place of articulation remain unchanged, movement is modulated.

Modulations for aspect tend to be changes in dynamic properties, such as rate, tension and acceleration, inviting description by such terms as thrust, tremulo, accelerando. Each modulation correlates with a grammatical category: predispositional, continuative, iterative, intensive, and so on. Often modulatory forms suggest their meaning, but their possible iconic origin does not interfere with their grammatical application. Thus, in the sign QUIET the hands move gently downward, but when its aspect is modulated by repetitive movement to mean "characteristically quiet," the hands move down in rapid, unquiet circles.

Once these inflectional processes had been discovered, whole sets of others came into view. ASL verbs are not inflected for tense: Time of occurrence is indexed for stretches of discourse, when necessary, by placing a sign along an arc from a point in front of the signer's face (future) to a point behind the ear (past). But ASL verbs are inflected for person, dual, number, reciprocal action and, using the same modulatory forms as adjectival predicates, for aspect.

As a step toward description of the system underlying inflectional structure, the authors posit eleven spatial and temporal dimensions of variation. The spatial dimensions include locus with respect to the three intersecting planes in front of the signer's torso, mentioned above, geometric pattern, and direction of movement; these dimensions are used to inflect for number and for the distribution of events over time, place, and participants in an action. The temporal dimensions include manner, rate, tension, evenness, and size of movement; these dimensions are used to inflect for manner, degree, and temporal aspect. Each dimension has only two or three values and many of the dimensions are independent, so that a single opposition often suffices to cue a distinction of meaning. A full featural account of ASL inflection may ultimately be possible, and the authors do, in fact, present a preliminary six-feature system that captures aspectual modulation of predicate adjectives.

The central puzzle, with which the authors leave us, is the relation between inflectional and lexical structure. The dimensions of movement that describe inflections are quite different from those that describe lexical forms. Often, the movements of uninflected signs seem to be embedded in the movement imposed by inflection, and indexical movements are superimposed on both. In other words, ASL appears to have three parallel formational systems:

lexical, morphological, and indexical. If this is really so, ASL differs radically from spoken languages where the same phonological segments are used for both lexical and morphological processes.

However, there is also evidence that this separation into layers may be more apparent than real. Supalla and Newport (1978) have shown that a lexical sign with repeated cycles of movement has only one cycle, when it is inflected for continuative aspect; similarly, a lexical sign with repeated downward movements loses all but one of them under modulation. Other signs with iterated, oscillating or wiggling movements in their surface lexical form are also reduced under modulation to a single base movement. And for yet other signs, lexical movement is not embedded in the modulation, but is transformed into a qualitatively different pattern. For such signs, at least, inflectional processes seem to operate not on the surface lexical form, but on an underlying stem. The authors conclude that a deeper analysis of ASL structure could reveal "...a unified internal organization which, in its systematicity, may bear a striking resemblance to equivalent levels of structure posited for spoken languages" (p. 315).

Whatever the outcome of this endeavor, the final chapters of Part III firmly establish ASL as an inflecting language, like Greek or Latin or Russian. They complete the demonstration that the dual structure of spoken language is not a mere consequence of mode, but a reflection of underlying cognitive structure. How far that cognitive structure was itself shaped by the (presumably) oral-auditory mode in which language evolved, we do not know. But language, as it now exists, can indeed be instantiated in another sensorimotor modality, and, when it is, its surface is shaped by properties of that modality.

What does this conclusion imply for the study of language and speech? Certainly not--and the authors firmly deny this inference--that speech is excluded from the biological foundations of language. Rather, we are impelled to study more closely the behavioral and neurological relations between vocal and manual articulation. The association between lateralizations for manual control and speech is well established. Recent studies have demonstrated that both skilled manual movements (Kimura & Archibald, 1974) and non-verbal oral movements (Mateer & Kimura, 1977) tend to be impaired in cases of nonfluent aphasia, and that disturbances of manual sign language in the deaf are associated with left hemisphere damage (Kimura, Battison, & Lubert, 1976). Evidence is also accumulating that sequential patterns of manual and vocal articulation are controlled by related neural centers (Kinsbourne & Hicks, 1979). Finally, preliminary studies at the Salk Institute (not reported in the present volume) have found behavioral evidence for left hemisphere superiority in the perception of ASL signs by native signers (Neville & Bellugi, 1978), suggesting the existence of a specialized sensorimotor mechanism, analogous to that for speech. The burden of all this work is that manual sign language belongs in the anatomical and physiological nexus of speech and language to which we alluded at the beginning of this review. The capacity for spoken and manual communication may rest on the evolution not only of the yet unformulated mechanisms that support abstract cognitive functions, but also of the fine, motor sequencing system in the left hemisphere by which those functions are expressed.

The discovery that language can be instantiated in another mode has implications for many other aspects of its study. Ultimately, language universals will have to be specified in a form general enough to capture the cognitive processes of both spoken and signed language. At present, the most fruitful study may be of language ontogeny. Logically, we still cannot exclude developmental mechanisms specialized for the discovery of language through speech. But the fact that deaf infants learn to sign, no less readily than their hearing peers learn to speak, argues for a broad adaptive mechanism, perhaps controlling the infant's search for patterned input in any communicatively viable modality (cf. Menn, 1979; Studdert-Kennedy, in press). The nature of this mechanism will surely be illuminated by comparisons between the ways deaf and hearing children learn their languages. Cross-linguistic studies are already under way at the Salk Institute and elsewhere. Indeed, the authors state in their introduction that the study of ASL acquisition was the initial impetus for the present work, and they promise a second volume reporting their developmental research.

Finally, as I look back on this splendid book, with its remorseless, subtle argument and its endless images of pert hands, winking and weaving, I am filled with admiration: for the deaf who invented the system of their extraordinary language, for the authors and their colleagues who are discovering it.

REFERENCE NOTE

1. Ganong, W. F. The internal structure of consonants in speech perception: Acoustic cues, not distinctive features. Unpublished manuscript.

REFERENCES

Aldridge, J. W. Levels of processing in speech perception. Journal of Experimental Psychology: Human Perception and Performance, 1978, 4, 164-177.

Baker, C., & Padden, C. A. Focusing on the nonmanual components of American Sign Language. In P. Siple (Ed.), Understanding language through sign language research. New York: Academic Press, 1978, 27-58.

Battison, R. Phonological deletion in American Sign Language. Sign Language Studies, 1974, 5, 1-19.

Campbell, R., & Dodd, B. Hearing by eye. Quarterly Journal of Experimental Psychology, in press.

Conrad, R. Speech and reading. In J. Kavanagh & I. Mattingly (Eds.), Language by ear and by eye. Cambridge, Mass.: MIT Press, 1972.

Du Brul, E. L. Origin of the speech apparatus and its reconstruction in fossils. Brain and Language, 1977, 4, 365-381.

Fischer, S. D. Sign language and Creoles. In P. Siple (Ed.), Understanding language through sign language research. New York: Academic Press, 1978, 309-332.

Friedman, L. A. (Ed.) On the other hand. New York: Academic Press, 1977.

Fromkin, V. A. The non-anomalous nature of anomalous utterances. Language, 1971, 47, 27-52.

Geschwind, N., & Levitsky, W. Human brain: Left-right asymmetries in temporal speech region. Science, 1968, 161, 186-187.

Grosjean, F. The perception of rate in spoken language and sign language.

Journal of Psycholinguistic Research, 1977, *22*, 408-413.

Kimura, D. The neural basis of language *qua* gesture. In H. Whitaker & H. A. Whitaker (Eds.), *Studies in neurolinguistics* (Vol. 3). New York: Academic Press, 1976.

Kimura, D., & Archibald, Y. Motor functions of the left hemisphere. *Brain*, 1974, *97*, 337-350.

Kimura, D., Battison, R., & Lubert, B. Impairment of nonlinguistic hand movements in a deaf aphasic. *Brain and Language*, 1976, *4*, 566-571.

Kroeber, A. L. Sign language inquiry. *International Journal of American Linguistics*, 1958, *24*, 1-19.

Kinsbourne, M., & Hicks, R. E. Mapping cerebral functional space: Competition and collaboration in human performance. In M. Kinsbourne (Ed.), *Asymmetrical function of the brain*. New York: Cambridge University Press, 1979, 267-273.

Lenneberg, E. H. *Biological foundations of language*. New York: Wiley, 1967.

Liberman, A. M., Cooper, F. S., Shankweiler, D. P., & Studdert-Kennedy, M. Perception of the speech code. *Psychological Review*, 1967, *74*, 431-461.

Liberman, A. M., & Studdert-Kennedy, M. Phonetic perception. In R. Held, H. W. Leibowitz, & H.-L. Teuber (Eds.), *Handbook of sensory physiology, Vol. VIII: Perception*. New York: Springer-Verlag, 1978, 143-187.

Lieberman, P. *The speech of primates*. The Hague: Mouton, 1972.

Liddell, S. K. Nonmanual signals and relative clauses in American Sign Language. In P. Siple (Ed.), *Understanding language through sign language research*. New York: Academic Press, 1978, 59-90.

Mateer, C., & Kimura, D. Impairment of non-verbal oral movements in aphasia. *Brain and Language*, 1977, *4*, 262-276.

Mayberry, R. I. Manual communication. In H. Davis & S. R. Silverman (Eds.), *Hearing and deafness* (4th ed.). New York: Holt, Rinehart & Winston, 1978.

Menn, L. *Pattern, control and contrast in beginning speech*. Bloomington: Indiana University Linguistics Club, 1979.

Miller, G. A., & Nicely, P. E. An analysis of perceptual confusions among English consonants. *Journal of the Acoustical Society of America*, 1958, *27*, 339-352.

Neville, H. J., & Bellugi, U. Patterns of cerebral specialization in congenitally deaf adults: A preliminary report. In P. Siple (Ed.), *Understanding language through sign language research*. New York: Academic Press, 1978, 239-260.

Parker, F. Distinctive features and acoustic cues. *Journal of the Acoustical Society of America*, 1977, *62*, 1051-1054.

Poizner, H., & Lane, H. Discrimination of location in American Sign Language. In P. Siple (Ed.), *Understanding language through sign language research*. New York: Academic Press, 1978, 271-288.

Siple, P. Visual constraints for sign language communication. *Sign Language Studies*, 1978, *19*.

Stokoe, W. C., Jr. Sign language structure. *Studies in Linguistics: Occasional Papers 8*. Buffalo: Buffalo University Press, 1960.

Stokoe, W. C., Jr. Classification and description of sign languages. In T. A. Sebeok (Ed.), *Current trends in linguistics* (Vol. XII). The Hague: Mouton, 1974, 345-371.

Stokoe, W. C., Jr., Casterline, D., & Croneberg, C. *A dictionary of American Sign Language on linguistic principles*. Washington, D.C.: Gallaudet College Press, 1965. (Second edition, 1976).

Studdert-Kennedy, M. The beginnings of speech. In G. B. Barlow, K. Immelmann, M. Main, & P. Petrinovich (Eds.), Behavioral development: The Bielefeld Interdisciplinary Project. New York: Cambridge University Press, in press.

Supalla, T., & Newport, E. L. How many seats in a chair? The derivation of nouns and verbs in American Sign Language. In P. Siple (Ed.), Understanding language through sign language research. New York: Academic Press, 1978, 91-132.

Umiker-Sebeok, D. J., & Sebeok, T. A. Aboriginal Sign 'Languages' from a semiotic point of view. Ars Semeiotica, 1977, 1, 69-97.

West, LaMont, Jr. The Sign Language: An analysis. Unpublished doctoral dissertation, Indiana University, Bloomington, 1960.

FOOTNOTES

[1] Eleven of the fourteen chapters were written in collaboration with one or more of the following: Robin Battison, Penny Boyes-Braem, Susan Fischer, Nancy Frishberg, Harlan Lane, Ella Mae Lentz, Don Newkirk, Elissa Newport, Carlene Canady Pederson, Patricia Siple.

[2] LaMont West, Jr.'s (1960) unpublished dissertation was an exception. At about the same time that Stokoe (1960) was beginning his analysis of ASL, West undertook to demonstrate, by morphemic and kinemic analysis, duality of patterning in Plains Sign Language (PSL). He isolated some eighty "kinemes," dividing them into five classes reminiscent of the Stokoe-Klima-Bellugi parameters of ASL: hand-shape, direction, motion-pattern, dynamics, and referent. West proposed parallels between kineme and phoneme classes, but was not fully satisfied by the parallels because of the large element of iconicity in PSL, and its tendency to form new signs with ad hoc handshapes which were not part of a closed kinemic system. West's work on PSL has not been followed up, but many of his doubts might be resolved by Klima and Bellugi's work on ASL.

[3] For fuller discussion than is appropriate here of errors commonly made in interpreting perceptual studies of speech sounds heard through noise, and of the distinction between linguistic features and their physical manifestations, see Parker (1977) and Ganong (Note 1).

[4] By convention, words in capital letters represent English glosses of ASL signs.

[5] Interestingly, West (1960) asserts of Plains Indian Sign Language that "...the obligatory grammatical relationships are established not by temporal order or syntax, but by spatial relationships..." and, further, that "...grammatical structure is almost entirely a matter of internal sign morphology..." (p. 90).

II. PUBLICATIONS

III. APPENDIX

PUBLICATIONS

Abramson, A. S. The phonetic plausibility of the segmentation of tones in Thai phonology. *Proceedings of the Twelfth International Congress of Linguists*. Innsbruck: Innsbrucker Beiträge zur Sprachwissenschaft, 1978, 760-763.

Abramson, A. S. The noncategorical perception of tone categories in Thai. In B. Lindblom & S. Ohman (Eds.), *Frontiers of speech communication research*. London: Academic Press, 1979.

Baer, T. Vocal jitter: A neuromuscular explanation. In V. Lawrence (Ed.), *Transcripts of the Eighth Symposium: Care of the professional voice*. New York: The Voice Foundation, in press.

Baer, T., Bell-Berti, F., & Tuller, B. On determining EMG onset time. In J. J. Wolf & D. H. Klatt (Eds.), *Speech communication papers presented at the 97th Meeting of the Acoustical Society of America*. New York: Acoustical Society of America, 1979, 75-78.

Borden, G. J. Use of feedback in established and developing speech. In N. J. Lass (Ed.), *Speech and language: Advances in basic research* (Vol. III). New York: Academic Press, in press.

Borden, G., & Harris, K. S. *Speech science primer: Physiology, acoustics, and perception of speech*. Baltimore: Williams and Wilkins, 1980.

Dorman, M. F., Raphael, L. J., & Liberman, A. M. Some experiments on the sound of silence in phonetic perception. *Journal of the Acoustical Society of America*, 1979, *65*, 1518-1532.

Feldman, L., & Turvey, M. T. Words written in Kana are named faster than the same words written in Kanji. *Language and Speech*, in press.

Fitch, H. L., Turvey, M. T., & Tuller, B. The Bernstein perspective: III. Tuning of coordinative structures with special reference to perception. In J. A. S. Kelso (Ed.), *Understanding human motor control*. Champaign, Ill.: Human Kinetics, in press.

Fowler, C. A. "Perceptual centers" in speech production and perception. *Perception & Psychophysics*, 1979, *25*, 375-388.

Fowler, C. A., Rubin, P., Remez, R., & Turvey, M. T. Implications for speech production of a general theory of action. In B. Butterworth (Ed.), *Language production*. New York: Academic Press, in press.

Fowler, C. A., & Turvey, M. T. On the concept of "command neuron" in explanations of behavior. *The Behavioral and Brain Sciences*, 1978, *1*, 20-22.

Fowler, C. A., & Turvey, M. T. Observational perspective and descriptive level in perceiving and acting. In W. Weimer & D. Palermo (Eds.), *Cognition and the symbolic processes II*. Hillsdale, N.J.: Erlbaum, in press.

Gay, T., & Turvey, M. T. Effects of efferent and afferent interference on speech production: Implications for a generative theory of speech motor control. *Proceedings of 9th International Congress of Phonetic Sciences*. Copenhagen: Institute of Phonetics, University of Copenhagen, 1979, 344-350.

Harris, K. S. Vowel duration change and its underlying physiological mechanisms. *Language and Speech*, 1978, *21*, 354-361.

Kelso, J. A. S. Motor-sensory feedback formulations: Are we asking the right questions? *The Behavioral and Brain Sciences*, 1979, *2*, 72-73.

Kelso, J. A. S. Introduction to the process approach to understanding skilled movement. In J. A. S. Kelso (Ed.), *Understanding human motor control*. Champaign, Ill.: Human Kinetics, in press.

Kelso, J. A. S. Overview of skilled performance: Coming to grips with the jargon. In J. A. S. Kelso (Ed.), *Understanding human motor control*. Champaign, Ill.: Human Kinetics, in press.

Kelso, J. A. S., Goodman, D., Hayes, C., & Stamm, C. F. Movement coding and memory in the developmentally young. *American Journal of Mental Deficiency*, 1979, *83*, 601-611.

Kelso, J. A. S., & Holt, K. G. Exploring a vibratory systems analysis of human movement production. *Journal of Neurophysiology*, in press.

Kelso, J. A. S., Holt, K. G., Kugler, P. N., & Turvey, M. T. On the concept of coordinative structures as dissipative structures. II. Empirical lines of convergence. In G. Stelmach & J. Requin (Eds.), *Tutorials in motor control*. Amsterdam: North Holland Publishing Co., in press.

Kelso, J. A. S., Pruitt, J. H., & Goodman, D. The anticipatory control of movement. In K. Newell & G. C. Roberts (Eds.), *Psychology of motor behavior and sport*. Champaign, Ill.: Human Kinetics, 1979, 192-204.

Kugler, P., Kelso, J. A. S., & Turvey, M. T. On the concept of coordinative structures as dissipative structures: I. Theoretical lines of convergence. In G. Stelmach & J. Requin (Eds.), *Tutorials in motor control*. Amsterdam: North Holland Publishing Co., in press.

Kugler, P. N., Kelso, J. A. S., & Turvey, M. T. On coordination and control in naturally developing systems. To appear in J. A. S. Kelso & J. C. Clark (Eds.), *The development of human movement control*. New York: John Wiley.

Kugler, P., & Turvey, M. T. Two metaphors for neural afference and efference. *The Behavioral and Brain Sciences*, in press.

Kuhn, G. M. Stop consonant place perception with single-formant stimuli: Evidence for the role of the front-cavity resonance. *Journal of the Acoustical Society of America*, 1979, *65*, 774-788.

Lee, W. A., & Kelso, J. A. S. Properties of slowly adapting joint receptors do not readily predict perception of limb position. *Journal of Human Movement Studies*, in press.

Löfqvist, A. Interarticulator programming in stop production. *Journal of Phonetics*, in press.

Lukatela, G., Popadić, D., Ognjenović, P., & Turvey, M. T. Lexical decision in a phonologically shallow orthography. *Memory & Cognition*, in press.

Lukatela, G., & Turvey, M. T. Some experiments on the Roman and Cyrillic alphabets of Serbo-Croatian. In R. Venesky & J. Kavanagh (Eds.), *Cross-cultural comparisons of orthographies, reading and dyslexia*. Baltimore, Md.: University Press, in press.

Mattingly, I. G. The skills of the plodder. (Review of *Toward a psychology of reading*, ed. by A. S. Reber & D. L. Scarborough. Hillsdale, N.J.: Erlbaum, 1977). *Contemporary Psychology*, 1978, *23*, 731-732.

Mattingly, I. G. Epimenides at the computer. (Review of *Gödel, Escher, Bach: The eternal golden braid*, by D. R. Hofstadter. New York: Basic Books, 1979). *Yale Review*, in press.

Mattingly, I. G. The psycholinguistic basis of linguistic awareness. *Twenty-eighth Yearbook of the National Reading Conference*, in press.

Michaels, C. F., & Turvey, M. T. Central sources of visual masking: Indexing structures supporting seeing at a single brief glance. Psychological Research, in press.

Miller, J. L., & Liberman, A. M. Some effects of later-occurring information on the perception of stop consonant and semivowel. Perception & Psychophysics, 1979, 25, 457-465.

Repp, B. H. Perceptual integration and differentiation of spectral cues for intervocalic stop consonants. Perception & Psychophysics, 1978, 24, 471-485.

Repp, B. H. Relative amplitude of aspiration noise as a voicing cue for syllable-initial stop consonants. Language and Speech, 1979, 22, 173-189.

Shaw, R., & Turvey, M. T. Coalitions as models for ecosystems: A realist perspective on perceptual organization. In M. Kubovy & J. Pomerantz (Eds.), Perceptual organization. Hillsdale, N.J.: Erlbaum, in press.

Shaw, R., & Turvey, M. T. Methodological realism. The Behavioral and Brain Sciences, in press.

Shaw, R., Turvey, M. T., & Mace, W. Ecological psychology: The consequence of a commitment to realism. In W. Weimer & D. Palermo (Eds.), Cognition and the symbolic processes II. Hillsdale, N.J.: Erlbaum, in press.

Streeter, L. A., & Nigro, G. N. The role of medial consonant transitions in word perception. Journal of the Acoustical Society of America, 1979, 65, 1533-1541.

Tuller, B., Fitch, H. L., & Turvey, M. T. The Bernstein perspective: II. The concept of muscle linkage or coordinative structure. In J. A. S. Kelso (Ed.), Understanding human motor control. Champaign, Ill.: Human Kinetics, in press.

Tuller, B., Harris, K. S., & Gross, B. Electromyographic study of the jaw muscles during speech. Journal of Phonetics, in press.

Turvey, M. T. The thesis of efference-mediation of vision cannot be rationalized. The Behavioral and Brain Sciences, 1979, 2, 81-83.

Turvey, M. T., & Remez, R. Visual control of locomotion in animals: An overview. In Proceedings of Conference on Interrelations among the communicative senses. NSF publication, 1979.

Turvey, M. T., & Shaw, R. The primacy of perceiving: An ecological reformation of perception for understanding memory. In L-G. Nilssen (Ed.), Perspectives in memory research: In honor of Uppsala University's 500th anniversary. Hillsdale, N.J.: Erlbaum, 1979.

Turvey, M. T., Tuller, B., & Fitch, H. L. The Bernstein perspective: I. The problems of degrees of freedom and context conditioned variability. In J. A. S. Kelso (Ed.), Understanding human motor control. Champaign, Ill.: Human Kinetics, in press.

Zimmerman, G., Kelso, J. A. S., & Lander, L. Articulatory behavior pre and post full-mouth tooth extraction and total alveoloplasty: A cinefluorographic study. Journal of Speech and Hearing Research, in press.

APPENDIX

DDC (Defense Documentation Center) and ERIC (Educational Resources Information Center) numbers:

Status Report		DDC	ERIC
SR-21/22	January - June 1970	AD 719382	ED-044-679
SR-23	July - September 1970	AD 723586	ED-052-654
SR-24	October - December 1970	AD 727616	ED-052-653
SR-25/26	January - June 1971	AD 730013	ED-056-560
SR-27	July - September 1971	AD 749339	ED-071-533
SR-28	October - December 1971	AD 742140	ED-061-837
SR-29/30	January - June 1972	AD 750001	ED-071-484
SR-31/32	July - December 1972	AD 757954	ED-077-285
SR-33	January - March 1973	AD 762373	ED-081-263
SR-34	April - June 1973	AD 766178	ED-081-295
SR-35/36	July - December 1973	AD 774799	ED-094-444
SR-37/38	January - June 1974	AD 783548	ED-094-445
SR-39/40	July - December 1974	AD A007342	ED-102-633
SR-41	January - March 1975	AD A013325	ED-109-722
SR-42/43	April - September 1975	AD A018369	ED-117-770
SR-44	October - December 1975	AD A023059	ED-119-273
SR-45/46	January - June 1976	AD A026196	ED-123-678
SR-47	July - September 1976	AD A031789	ED-128-870
SR-48	October - December 1976	AD A036735	ED-135-028
SR-49	January - March 1977	AD A041460	ED-141-864
SR-50	April - June 1977	AD A044820	ED-144-138
SR-51/52	July - December 1977	AD A049215	ED-147-892
SR-53	January - March 1978	AD A055853	ED-155-760
SR-54	April - June 1978	AD A067070	ED-161-096
SR-55/56	July - December 1978	AD A065575	ED-166-757
SR-57	January - March 1979	**	ED-170-823
SR-58	April - June 1979	**	ED-178-967
SR-59/60	July - December 1979	**	**

AD numbers may be ordered from:

U.S. Department of Commerce
National Technical Information Service
5285 Port Royal Road
Springfield, Virginia 22151

ED numbers may be ordered from:

ERIC Document Reproduction Service
Computer Microfilm International
Corp. (CMIC)
P.O. Box 190
Arlington, Virginia 22210

Haskins Laboratories Status Report on Speech Research is abstracted in Language and Behavior Abstracts, P.O. Box 22206, San Diego, California 92122.

**DDC and/or ERIC order numbers not yet assigned.

UNCLASSIFIED
Security Classification

DOCUMENT CONTROL DATA - R & D

(Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified)

1. ORIGINATING ACTIVITY (Corporate author)

Haskins Laboratories, Inc.
270 Crown Street
New Haven, Connecticut 06510

2a. REPORT SECURITY CLASSIFICATION

UNCLASSIFIED

2b. GROUP

N/A

3. REPORT TITLE

Haskins Laboratories Status Report on Speech Research, No. 59/60, July-December, 1979.

4. DESCRIPTIVE NOTES (Type of report and, inclusive dates)

Interim Scientific Report

5. AUTHOR(S) (First name, middle initial, last name)

Staff of Haskins Laboratories, Alvin M. Liberman, P.I.

6. REPORT DATE

December, 1979

7a. TOTAL NO. OF PAGES

274

7b. NO. OF REFS

557

8a. CONTRACT OR GRANT NO.

HD-01994	NS13870
N01-HD-1-2420	NS13617
RR-05596	AM25814
BNS76-82023	
MCS76-81034	
MCS79-16177	

9a. ORIGINATOR'S REPORT NUMBER(S)

[illegible]-59/60 (1979)

9b. OTHER REPORT NO(S) (Any other numbers that may be assigned this report)

None

10. DISTRIBUTION STATEMENT

Distribution of this document is unlimited*

11. SUPPLEMENTARY NOTES

N/A

12. SPONSORING MILITARY ACTIVITY

See No. 8

13. ABSTRACT

This report (1 July - 31 December) is one of a regular series on the status and progress of studies on the nature of speech, instrumentation for its investigation, and practical applications. Manuscripts cover the following topics:

See Table of Contents.

DD FORM 1 NOV 65 1473 (PAGE 1)

S/N 0101-807-6811

*This document contains no information not freely available to the general public. It is distributed primarily for library use.

UNCLASSIFIED
Security Classification

A-31408

14 KEY WORDS	LINK A		LINK B		LINK C	
	ROLE	WT	ROLE	WT	ROLE	WT
Articulation						
Laryngeal activity						
in consonant clusters, geminates						
Prosody, sentence and lexical (Thai)						
Vocal tract length, preservation						
Perception						
Brain lateralization in infants						
Discrimination of phonetics and musical timbre						
Consonants						
Stops, Influence of preceding fricative						
Fricatives, Influence of vocalic context						
Silence, Influence of vocalic environment						
Stimulus dominance						
Reading						
Sentence memory, effect on reading ability						
Adaptation						
Effects between song and string						
Sign Language						
Book review						
Computer						
Book review						